2004

Recreation demand: on-site sampling and responsiveness of trip behavior to physical water quality measures

Kevin Joseph Egan
Iowa State University

Recreation demand: On-site sampling and responsiveness of trip behavior to physical water quality measures

by

Kevin Joseph Egan

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Economics

Program of Study Committee:
Joseph A. Herriges, Co-major Professor
Catherine L. Kling, Co-major Professor
Arne Hallam
Jean-Didier Opsomer
Jinhua Zhao

Iowa State University

Ames, Iowa

2004

Graduate College

Iowa State University

This is to certify that the doctoral dissertation of

Kevin Joseph Egan

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Co-major Professor

Signature was redacted for privacy.

Co-major Professor

Signature was redacted for privacy.

For the Major Program

To my wife, Stephanie, for all her love and support.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I thank my wife, Stephanie, and her parents, Stephen and Nancy, for their encouragement and understanding throughout my years of graduate school. I also thank my parents, John and Debbie, for creating an environment where education is important.

Lastly, I thank my major professors, Joe Herriges and Cathy Kling, who offered immeasurable guidance.

CHAPTER 1. GENERAL INTRODUCTION

I. Introduction

The first two essays of this dissertation focus on one lake in particular, Clear Lake, located in Clear Lake, Iowa. Clear Lake was formed by glacial action during the last ice age and is the third largest natural lake in Iowa. In 1950, Clear Lake was pristine. The water was clear blue with an object being visible 5 to 8 feet under the surface. However, in the last 50 years the lake has deteriorated to a level that causes concern. Today, visibility is about 6 inches to one foot. Other water quality measures have also deteriorated, for example, Clear Lake now experiences occasional algae blooms and a decreased diversity of fish populations.

Despite the deteriorated conditions of the lake, Clear Lake is still the center of many activities, and is especially lively in the summer months. Anglers, recreational boaters, sailors, and beach users all frequent the lake. It is a valuable resource to the city of Clear Lake and the state of Iowa, generating over $30 million a year in tourism revenues (Downing and Kopaska, 2001). If preserved, Clear Lake will remain one of Iowa's unique destinations for recreationists.

Valuing preservation and the improvement of water quality at Clear Lake was the purpose of a survey mailed to visitors and local residents in the summer of 2000. The first two essays in this dissertation focus on the visitors' survey and different ways to use the information the respondents provide. Specifically, the Clear Lake survey asks revealed preference (RP) and stated preference (SP) questions. The first two essays combine this RP and SP data with the difference between the two being the particular SP information used from the survey.

In the first essay, contingent behavior trips which are contingent on price (i.e. travel cost) changes presented in the survey are the SP data that is combined with the RP data. One focus is to pool the RP and SP data to better estimate an average demand curve used in the travel cost model. However, since the Clear Lake data was gathered by intercepting the visitors on-site, the primary purpose of the first essay is to consider the problem of controlling for on-site sampling in the context of a panel (i.e. the visitors RP and SP responses) of demand equations. This essay is the first to address controlling for on-site sampling with panel data. A multivariate Poisson-log normal model is used to jointly model the RP and SP data and to correct for on-site sampling.

In the second essay the objective is a combined RP and SP model to estimate willingness to pay (WTP) for water quality improvements. Therefore, the SP data used is contingent behavior trips contingent on water quality changes and also contingent valuation data about the same water quality scenarios. A continuous model is utilized that can exploit the economic theory of consumers. This theoretically consistent model jointly estimates the above three data sources, one RP and two SP (i.e. the contingent behavior and contingent valuation data). Again, since the data is collected on-site this model is corrected for on-site sampling.

The third essay utilizes a more recent data set, the first year survey of the Iowa lakes project mailed in the fall of 2002. This random population survey was sent to 8,000 Iowans collecting information on their recreation behavior to 129 lakes. The lakes were chosen to coincide with the research being done by the Iowa State University Limnology Laboratory led by John Downing, an ISU limnologist and professor. He is commissioned to collect numerous physical water quality measures three times per year for 5 years at each of these

129 lakes. Such a complete record of water quality will be combined with the results from the recreation demand survey.

The objective of this essay is to analyze Iowan's recreation behavior to the 129 principle lakes and their responsiveness to physical water quality measures. It is expected that lakes with excessive nutrients will therefore have more algae blooms, decreased water clarity, and undesirable color and odor that will lead to these lakes being less visited. The results do confirm this hypothesis and even more allow policy relevant compensating variation scenarios based on improvements in the lakes physical water quality measures. A random utility model, specifically a repeated mixed logit, is employed to analyze the individual's trip behavior to the 129 lakes.

To conclude, in all three essays recreation demand modeling is the primary tool. In the first two essays stated preference data is combined with the revealed preference data. The stated preference (contingent behavior) data is used in the first essay to obtain a more complete picture of the visitor's responsiveness to travel costs, and in the second essay (contingent valuation data) to ask for the visitor's willingness to pay given proposed water quality improvements. The third essay exclusively utilizes revealed preference data as the 129 lakes offer variation in water quality unlike few other places. Many lakes contain nutrient levels that are some of the highest in the world while others compare to Minnesota's northern lakes in terms of overall water quality.

II. Dissertation Organization

Each of the three essays in this dissertation is a separate chapter, with its own introduction, conclusion, and references. A general conclusion chapter summarizes the results from all three essays. Finally, two appendixes are included. The first appendix is the

Clear Lake Visitors Survey used in the first two essays, and the second appendix is the first year Iowa Lakes Survey used in the third essay.

III. References

Downing, John A., and Jeff Kopaska (2001). "Summary of the Clear Lake Diagnostic Report."

CHAPTER 2. MIXED POISSON REGRESSION MODELS WITH INDIVIDUAL PANEL DATA FROM AN ON-SITE SAMPLE

A paper submitted to the Journal of Environmental Economics and Management

Kevin J. Egan[1,2] and Joseph A. Herriges[1,3]

I. Introduction

Cost considerations often drive analysts to rely upon intercept (or on-site) surveys to collect information about recreation demand at a site (or sites) of interest. This guarantees that survey respondents will include users of the resource in question. Unfortunately, the sampling procedure also comes at a cost of both truncation (excluding non-users) and endogenous stratification (over sampling those individuals who are more frequent users of the site). As a result, the sample is no longer representative of the broader population. Failure to correct for on-site sampling will result in biased estimates of recreation demand and any corresponding welfare estimates.

There have been a number of papers in the literature focused on controlling for intercept sampling in recreation demand analysis. Shaw (1988) develops a correction for both the truncation and endogenous stratification problems in the case of a single site Poisson count data model. Englin and Shonkwiler (1995) subsequently extended Shaw's correction to the case of the Negative Binomial (NB) count data model. The advantage of the NB model is that it allows for overdispersion (i.e., the situation in which the conditional mean number of trips is less than the conditional variance of trips), a common characteristic of recreation

[1] Predoctoral research associate and Professor, respectively, Department of Economics, Iowa State University.
[2] Primary researcher.
[3] Author for correspondence.

demand data. The limitation of both of these efforts is that they are focused on a single demand equation.

The purpose of this paper is to consider the problem of controlling for on-site sampling in the context of a system (or panel) of demand equations.[4] Specifically, we are concerned with the situation in which survey respondents are asked to provide information not only about the actual trips to a specific site (observed behavior), but also their anticipated trips (either under current conditions or given price and quality changes). The latter trip data, typically known as contingent behavior data, has been used to study the impact of changing environmental conditions (See, e.g., Rosenberger and Loomis, 1999; Whitehead *et al.*, 2000; and Grijalva *et al.*, 2002). Unfortunately, if the observed and contingent behavior data are collected through an on-site survey, the sampling problems become more complex. The observed behavior data are, as before, subject to truncation and endogenous stratification. While the contingent behavior data are not directly impacted, they are *incidentally* truncated and endogenously stratified. That is, while the sampling does not exclude individuals who anticipate zero trips in the future, they are less likely because the sampling procedure has excluded individuals who took zero trips in the past and oversampled individuals who, at least in the past, frequently took trips. As a result, it is important to model the observed and contingent behavior data in a panel data framework, controlling for correlation between these data sources and the sampling mechanism used.

In this paper, the multivariate Poisson-log normal (MPLN) model is used to jointly model the observed and contingent behavior data and to correct for on-site sampling.

---

[4] The literature has already shown a need for this research as evidenced by Englin et al. (2001), who acknowledge their inability to estimate population values since their panel data was collected on-site.

Aitchison and Ho (1989) first suggested the MPLN model but did not include regressors in their analysis. Munkin and Trivedi (1999) estimate a bivariate PLN model. The advantage of the MPLN specification is the fact that, as Shonkwiler (1995) notes, "...only the multivariate Poisson-lognormal distribution can both reproduce an arbitrary correlation structure and account for overdispersion." We modify the MPLN model to control for on-site sampling.

The resulting model is used to analyze survey data collected on-site at Clear Lake in north central Iowa. Specifically, the survey data included observed trips for 2000 and contingent behavior trips for 2001 under both current prices and two sets of price increases. We find a substantial bias results if the sampling procedures are ignored, overstating both the average number of trips to the site (by a factor of 11) and the welfares associated with the recreational opportunities at Clear Lake.

II. Correcting for On-Site Sampling

It has long been recognized that, while on-site (or intercept) surveys provide a convenient mechanism for insuring that a sample includes site users, the resulting sample is no longer representative of the population as a whole. This section provides an overview of the corrections developed for the single-site setting. These corrections are then extended for the multivariate scenario.

A. The Univariate Model

Shaw (1988) was the first to recognize the complex set of problems that characterize on-site samples in recreation demand analysis. In addition to the count nature of the data (i.e., non-negative integers), he notes that on-site surveys exclude those who do not visit the site (truncation) and over sample those who frequent the site regularly (endogenous stratification).[5] His correction for these problems, based on the Poisson regression model, is both intuitive and easy to implement.[6]

Shaw (1988) begins by assuming that population trips to the single site of interest follow a univariate Poisson distribution. That is,

$$f(y_i \mid x_i) = \frac{\exp(-\lambda_i)(\lambda_i)^{y_i}}{y_i!}, \quad y_i = 0,1,2,\ldots \tag{1}$$

where y_i denotes the number of trips taken by individual i,

$$\begin{aligned} \lambda_i &= E(y_i \mid x_i) \\ &= \exp(\beta' x_i) \end{aligned} \tag{2}$$

denotes the expected number of trips for an individual with characteristics vector x_i, and β denotes the unknown parameters of the distribution to be estimated.

In correcting for the on-site sampling, Shaw assumes that visitors taking y_i trips are y_i times more likely to be sampled than someone who takes only one trip. He demonstrates that the on-site sample's distribution is then the product of the population distribution and odds (relative to an average individual) of being included in the sample; i.e.,

$$\begin{aligned} f_{OS}(y_i \mid x_i) &= \frac{y_i}{E(y_i \mid x_i)} f(y_i \mid x_i) \\ &= \frac{y_i}{\lambda_i} \frac{\exp(-\lambda_i)(\lambda_i)^{y_i}}{y_i!} \\ &= \frac{\exp(-\lambda_i)(\lambda_i)^{y_i-1}}{(y_i-1)!}, \quad y_i = 1,2,\ldots \end{aligned} \tag{3}$$

---

[5] As Shaw (1988) notes, a number of authors recognized earlier the truncation issue associated with on-site surveys, including Smith and Desvousges (1985). The issue of truncation in recreation demand was further discussed by Creel and Loomis (1990) and Grogger and Carson (1991).

[6] Shaw (1988) actually provides two solutions to the on-site sampling, one based on the Poisson regression model and a second based on a continuous regression model of trip data. We focus our attention here on the count data model, though the corrections could be adapted for the continuous setting.

The form of the on-site sample's distribution is convenient since it can be estimated using standard statistical packages designed to estimate a Poisson regression model. The only change required for on-site sampling is to replace y_i with $y_i - 1$ as the dependent variable.

One limitation of Shaw's model is, like all Poisson models, it imposes the assumption of equidispersion; i.e.,

$$\lambda_i = E(y \mid x_i) = Var(y_i \mid x_i). \tag{4}$$

In practice, however, recreation demand data typically exhibit overdispersion with the conditional trip variance exceeding the conditional trip mean. Following the logic of Shaw, Englin and Shonkwiler (1995) extend the on-site corrections to the negative binomial model. Specifically, if population trips are characterized by the negative binomial distribution

$$f(y_i \mid x_i) = \frac{\Gamma\left(y_i + \alpha_i^{-1}\right)\alpha_i^{y_i}\lambda_i^{y_i}\left(1 + \alpha_i\lambda_i\right)^{-\left(y_i + \alpha_i^{-1}\right)}}{\Gamma\left(y_i + 1\right)\Gamma\left(\alpha_i^{-1}\right)}, \tag{5}$$

then the on-site sample will be characterized by the distribution

$$f_{OS}(y_i \mid x_i) = \frac{y_i\Gamma\left(y_i + \alpha_i^{-1}\right)\alpha_i^{y_i}\lambda_i^{y_i - 1}\left(1 + \alpha_i\lambda_i\right)^{-\left(y_i + \alpha_i^{-1}\right)}}{\Gamma\left(y_i + 1\right)\Gamma\left(\alpha_i^{-1}\right)}. \tag{6}$$

In this case the mean and variance for the on-site sample are

$$E(y_i \mid x_i) = \lambda_i + 1 + \alpha_i\lambda_i \tag{7}$$

and

$$Var(y_i \mid x_i) = \lambda_i\left(1 + \alpha_i + \alpha_i\lambda_i + \alpha_i^2\lambda_i\right), \tag{8}$$

allowing for overdispersion and reducing to Shaw's Poisson model when $\alpha_i \rightarrow 0$.

B. The Multivariate Setting

The results of the previous section apply only to the univariate setting. However, there are many examples in practice where a system of counts must be modeled. This is the case, for example, if intercept surveys are conducted at several sites simultaneously or if trip data are gathered at a single site for a series of years or under a series of hypothetical or actual scenarios. Laitila (1999) has addressed the former problem using independent Poisson distributions for each site and conditioning on the total number of trips taken. In this paper, we focus our attention on the latter problem. As noted above, the latter scenario has arisen in recent years, as recreation demand surveys frequently ask not only for information on past trips (observed behavior), but also inquire as to changes in trip behavior in future years and under hypothetical changes to the recreation site of interest (contingent behavior). We begin this section by reviewing the multivariate count data models and then develop corrections to those models for on-site samples.

1. Multivariate Count Data Models

The simplest extension of the univariate Poisson count data model to the multivariate setting is to assume that trip data follow independent Poisson distributions. Specifically, if y_{ij} denotes the number of trips that individual *i* would take (or has taken) under scenario *j*, then the joint conditional distribution for the vector of trips $y_{i\cdot} = \left(y_{i1}, \ldots, y_{iJ}\right)'$ is given by

$$f\left(y_{i\cdot} \mid x_{i\cdot}\right) = \prod_{j=1}^{J} \frac{\exp\left(-\lambda_{ij}\right)\left(\lambda_{ij}\right)^{y_{ij}}}{y_{ij}!}, \quad y_{ij} = 0,1,2,\ldots \tag{9}$$

where

$$\begin{aligned}\lambda_{ij} &= E\left(y_{ij} \mid x_{ij}\right) \\ &= \exp\left(\beta_j' x_{ij}\right)\end{aligned} \tag{10}$$

and $x_{i\cdot} = \left(x_{i1}, \ldots, x_{iJ}\right)'$.

The problem with the model in (9) is that the assumption of independence is unlikely to hold in practice. Individuals who have taken a large number of trips in the past (say y_{i1}) are also likely to take a large number of trips in the future or under proposed changes to the site being studied (i.e., $y_{i2}, \ldots, y_{iJ}$). There have been a number of multivariate count data models developed in the literature to allow for correlation across counts for the same individual. Most of these models are mixed Poisson specifications that allow for a common shared source of unobserved heterogeneity in the counts for a given individual. Mixed Poisson models begin by assuming that there is an unobserved factor, $v_{ij} = \exp\left(\varepsilon_{ij}\right)$, associated with trips taken by individual i under scenario j. If v_{ij} were known, then the corresponding trips would follow a standard Poisson process, with

$$f\left(y_{ij} \mid x_{ij}, v_{ij}\right) = \frac{\exp\left(-\tilde{\lambda}_{ij}\right)\left(\tilde{\lambda}_{ij}\right)^{y_{ij}}}{y_{ij}!}, \quad y_{ij} = 0, 1, 2, \ldots \tag{11}$$

and

$$\begin{aligned}E\left(y_{ij} \mid x_{ij}, v_{ij}\right) &= \tilde{\lambda}_{ij} \\ &= \lambda_{ij} v_{ij} \\ &= \exp\left(\beta_j' x_{ij} + \varepsilon_{ij}\right),\end{aligned} \tag{12}$$

With the v_{ij} (or equivalently ε_{ij}) being unobserved, the relevant distribution for $y_{i\cdot}$ becomes

$$f(y_{i\cdot} \mid x_{i\cdot}) = \int \cdots \int \prod_{j=1}^{J} \frac{\exp\left(-\lambda_{ij} \exp(\varepsilon_{ij})\right)\left(\lambda_{ij} \exp(\varepsilon_{ij})\right)^{y_{ij}}}{y_{ij}!} g(\varepsilon_{i\cdot}) d\varepsilon_{i1} \cdots d\varepsilon_{iJ}, \quad y_{ij} = 0,1,2,\ldots \quad (13)$$

where $g(\varepsilon_{i\cdot})$ denotes the pdf for $\varepsilon_{i\cdot}$. Thus, the distribution of the trip vector, $y_{i\cdot}$, becomes a mixture of Poisson distributions. There are two consequences of this mixing process. First, the equidispersion assumption in equation (4) will no longer apply to the individual trip data (i.e., the y_{ij}'s). Second, allowing for correlation among the ε_{ij}'s across scenarios (*j*) for a given individual (*i*) will induce correlation among the corresponding y_{ij}'s for that individual.

In this paper, we will focus our attention on one such mixed Multivariate Poisson model, the Multivariate Poisson-Lognormal distribution (MPLN).[7] The MPLN model was introduced by Aitchison and Ho (1989) and gets its name from the fact that the vector $v_{i\cdot}$ is assumed to follow a multivariate lognormal distribution, or equivalently that $\varepsilon_{i\cdot}$ follows a multivariate normal distribution; i.e.,

$$\varepsilon_{i\cdot} \sim N(0, \Omega). \quad (14)$$

Substituting this distributional assumption into (13), we then have that

$$f(y_{i\cdot} \mid x_{i\cdot}) = \int \cdots \int \prod_{j=1}^{J} \frac{\exp(-\tilde{\lambda}_{ij})(\tilde{\lambda}_{ij})^{y_{ij}}}{y_{ij}!} \frac{\exp\left[-\frac{1}{2}\varepsilon_{i\cdot}'\Omega^{-1}\varepsilon_{i\cdot}\right]}{(2\pi)^{J/2}|\Omega|^{1/2}} d\varepsilon_{i\cdot}, \quad y_{ij} = 0,1,2,\ldots \quad (15)$$

The conditional trip means and variances become

$$E\left[y_{ij} \mid x_{ij}\right] = \lambda_{ij} \exp\left(\tfrac{1}{2}\sigma_j^2\right) \equiv \delta_{ij} \quad (16)$$

and

[7] The MPLN model can be viewed as incorporating random individual effects. An alternative approach would be to allow for individual fixed effects. Hausman, Hall and Griliches (1984) develop a fixed effects model in the context of patents and R&D expenditures. Englin and Cameron (1996) apply their model in the recreation demand context.

$$Var\left[y_{ij} \mid x_{ij}\right] = \delta_{ij} + \left[\exp\left(\sigma_j^2\right) - 1\right]\delta_{ij}^2, \tag{17}$$

where $\sigma_j^2 = Var\left(\varepsilon_{ij} \mid x_{ij}\right)$. Thus, equidispersion results only if $\sigma_j \to 0$. Correlation among the trips emerges because

$$Cov\left[y_{ij}, y_{ik}\right] = \delta_{ij}\left[\exp\left(\sigma_{jk}\right) - 1\right]\delta_{ik},\ j \neq k. \tag{18}$$

where σ_{jk} denotes the $(j,k)^{th}$ element of Ω. One of the attractive features of the MPLN specification is that it does not restrict the sign of this correlation. The correlation between trips for two distinct scenarios *j* and *k* can be positive, negative, or zero and depends directly upon the sign of the corresponding σ_{jk}. The downside of the MPLN specification is that, at the estimation stage, the pdf in (15) requires integration over a *J*-dimensional integral. However, either standard numerical procedures or simulation techniques can be used to address this problem as long as the number of scenarios, *J*, remains relatively small; i.e., less than eight.

An alternative to the MPLN model is the Multivariate Poisson Gamma (MPG) specification.[8] In this case, it is assumed that there is a single unobserved factor, u_i, shared by all trip scenarios for the same individual; i.e.,

$$v_{ij} = u_i \ \forall j \tag{19}$$

and that u_i follows a $gamma(\alpha, \alpha)$ distribution with a mean of 1 and a variance of α^{-1}. Substituting this assumption into (13) yields[9]

---

[8] The MPG specification was introduced by Arbous and Kerrich (1951) in a bivariate context and subsequently extended by Bates and Neyman (1952) and Nelson (1985). In the economics literature, Hausman, Hall and Griliches (1984) use the MPG model as a random effects model to capture correlation between patents and R&D expenditures.

[9] See Winkelmann (2000, p. 196).

$$f\left(y_{i\cdot} \mid x_{i\cdot}\right)=\frac{\Gamma\left(\sum_{j=1}^{J} y_{ij}+\alpha\right)\alpha^{\alpha}\left(\sum_{j=1}^{J}\lambda_{ij}+\alpha\right)^{-\left(\sum_{j=1}^{J} y_{ij}+\alpha\right)}}{\Gamma(\alpha)}\prod_{j=1}^{J}\frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!},\quad y_{ij}=0,1,2,\ldots \tag{20}$$

The corresponding conditional means and variances are given by

$$E\left[y_{ij} \mid x_{ij}\right]=\lambda_{ij} \tag{21}$$

and

$$V\left(y_{ij} \mid x_{ij}\right)=\lambda_{ij}+\alpha^{-1}\left(\lambda_{ij}\right)^{2}. \tag{22}$$

Thus, the degree of overdispersion is a decreasing function of α. The covariance between trip responses for a given individual becomes

$$Cov\left[y_{ij}, y_{ik}\right]=\alpha^{-1}\lambda_{ij}\lambda_{ik}. \tag{23}$$

One advantage of the MPG specification is the closed form nature of the count probabilities in equation (20), avoiding the need for numerical or simulation based integration when estimating the model. However, unlike the MPLN, the MPG imposes considerable structure on the correlation among the counts, requiring the correlations to always be positive and driven by the single parameter α.

2. Controlling for On-Site Sampling

The problem of on-site sampling emerges for the application we are considering because the first of the trip scenarios, j=1, corresponds to current trips to the site in question. Thus, y_{i1} is truncated, excluding observations in the population with $y_{i1}=0$, and endogenously stratified, with the sample over representing individuals that frequently visit the site. If we were only interested in observed trip behavior, then the univariate Poisson, Negative Binomial (both described in the previous section), or the univariate PLN model could be applied. However, individuals visiting the site are asked not only about their actual trip taking behavior to the site, but also about how often they plan to visit the site in future years and under a variety of possible changes to the site, generating a vector of trip counts $y_{i\cdot} = (y_{i1}, y_{i2}, \ldots, y_{iJ})'$. The contingent behavior trips $y_{i,-1} \equiv (y_{i2}, \ldots, y_{iJ})'$, while not directly truncated or endogenously stratified, are impacted by the on-site nature of the survey through the correlation between y_{i1} and $y_{i,-1}$. Specifically, following the same logic as Shaw (1988) used in the univariate case,

$$f_{OS1}(y_{i\cdot} \mid x_{i\cdot}) = \frac{y_{i1}}{E(y_{i1} \mid x_{i\cdot})} f(y_{i\cdot} \mid x_{i\cdot}), \quad y_{i1} = 1, 2, \ldots; y_{i,-1} = 0, 1, \ldots \tag{24}$$

where the subscript *OS1* is used to denote the fact that the on-site sampling *directly* impacts the trips for scenario *j*=1.

If the trips are independently distributed and each follow a Poisson process, then

$$f_{OS1}(y_{i\cdot} \mid x_{i\cdot}) = \frac{\exp(-\lambda_{i1})(\lambda_{i1})^{y_{i1}-1}}{(y_{i1}-1)!} \prod_{j=2}^{J} \frac{\exp(-\lambda_{ij})(\lambda_{ij})^{y_{ij}}}{y_{ij}!}, \quad y_{i1} = 1, 2, \ldots; y_{i,-1} = 0, 1, \ldots \tag{25}$$

If the MPLN specification applies, however, then

$$f_{OS1}(y_{i\cdot} \mid x_{i\cdot}) = \int \cdots \int \frac{y_{i1}}{\delta_{i1}} \prod_{j=1}^{J} \frac{\exp(-\tilde{\lambda}_{ij})(\tilde{\lambda}_{ij})^{y_{ij}}}{y_{ij}!} \frac{\exp\left[-\frac{1}{2}\varepsilon_{i\cdot}'\Omega^{-1}\varepsilon_{i\cdot}\right]}{(2\pi)^{J/2}|\Omega|^{1/2}} d\varepsilon_{i\cdot}, \quad y_{i1} = 1, 2, \ldots; \quad y_{i,-1} = 0, 1, \ldots \tag{26}$$

A similar correction applies for the MPG specification, yielding:

$$f_{OS1}\left(y_{i\cdot} \mid x_{i\cdot}\right)=\frac{y_{i1}\Gamma\left(\sum_{j=1}^{J} y_{ij}+\alpha\right)\alpha^{\alpha}\left(\sum_{j=1}^{J}\lambda_{ij}+\alpha\right)^{-\left(\sum_{j=1}^{J} y_{ij}+\alpha\right)}}{\lambda_{i1}\Gamma(\alpha)}\prod_{j=1}^{J}\frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!},\quad y_{i1}=1,2,\ldots;\quad y_{i,-1}=0,1,\ldots \tag{27}$$

While we have estimated the MPG model, the results were clearly dominated by the MPLN specification in terms of a likelihood dominance criterion and the Akaike information criterion. In the remainder of this paper, we focus our attention exclusively on the MPLN specification, though the results from the MPG model are available from the authors upon request.

III. Data and Model Specification

The data used in our empirical application are drawn from an intercept survey of visitors to Clear Lake located in north central Iowa. Visitors' names and addresses were collected on-site in the summer of 2000. These individuals were then mailed a survey in October, 2000. The survey asked respondents to provide four trip totals:

- Observed Behavior (OB): Their total number of trips to Clear Lake between November 1999 and October 2000.
- Contingent Behavior (CB_0): Their anticipated number of trips in 2001, given current travel costs.
- Contingent Behavior (CB_1): Their anticipated number of trips in 2001, given an increase in the total cost per trip of $\$B$. Specifically, individuals were asked: "Suppose that the price of visiting Clear Lake increases by $\$B$ per trip (due for example to gas prices, user fees, or equipment costs). How many times would you

visit next year?" The value of B was randomly assigned to each survey respondent and varied across individuals in the sample from \$3 to \$15, with a mean of \$7.26.

- Contingent Behavior (CB_2): Their anticipated number of trips in 2001, given a price increase of \$$C$ per trip, where C>B. Again, the value of C was randomly assigned to each survey respondent and varied across individuals in the sample from \$7 to \$30, with a mean of \$16.88.

In addition to gathering trip data, the survey also asked a series of contingent valuation questions, inquired as to the respondents' attitudes towards water quality improvements, and gathered socio-demographic information.

Of the 1,024 individuals intercepted at Clear Lake, 626 (or 62.7% of the deliverable surveys) returned a completed mail survey. In the analysis below, individuals were excluded from the final sample if they reported seasonal trips in excess of 52, allowing one trip per weekend. This resulted in 36 individuals being excluded from the sample. We also excluded households whose travel time was greater than five hours one way. Clear Lake is a unique natural lake in Iowa and does draw travelers from around the state. However, it is a regional attraction and the assumption is that anyone traveling from farther than five hours likely made the journey primarily for reasons other than to visit the lake. This excluded 19 additional households. Finally, for simplicity, a balanced panel was obtained by excluding visitors who did not answer all of the trip questions. The final sample size used in the analysis was N=543.

In the models estimated below, the average number of trips under scenario j (λ_{ij})is assumed to be a function of the travel cost to Clear Lake, household income, and socio-demographic characteristics of the household. Specifically,

$$\lambda_{ij} = \exp\left(\beta_{0j} + \beta_{Pj}P_{ij} + \beta_{Ij}I_i + \delta_j' z_i\right), \tag{28}$$

where P_{ij} denotes the roundtrip travel costs from individual *i*'s home to Clear Lake and back, I_i denotes individual *i*'s annual income, and z_i is a vector of socio-demographic characteristics of the household, including:

- *Male* =1 if the survey respondent is male, =0 otherwise;
- *Age* = the age of the survey respondent;
- Age^2;
- *School* = 1 if the survey respondent has attended or completed some level of post-high school education; and
- *Household* = the total number of household members.

For observed trips (OB) and forecasted trips for 2001 (CB_0), travel costs were computed as \$0.25 times the round-trip travel distance, computed using *PCMiler*, plus one third the respondent's wage rate times their round-trip travel time. P_{ij} for CB_1 and CB_2 are computed in the same fashion, except that \$B and \$C are added to the travel costs, respectively.

Table 1 provides a summary of the data used in the analysis. There are a number of attributes of the raw trip data that are worth noting. First, for all four trip variables, the unconditional mean number of trips in the sample is roughly the same order of magnitude as the corresponding unconditional standard deviation, indicating that the unconditional variance will be eight to twelve times the unconditional mean. This suggests that overdispersion is likely to be a problem for all four trip variables and that a simple Poisson model for each trip variable will be inappropriate. Second, the observed number of trips (OB) is large, with households in the sample averaging over a dozen trips per year. This should not, however, be interpreted as indicative of the population as a whole, but rather a reflection of the on-site sampling process. Households who frequent Clear Lake are more likely to be included in the sample precisely because they were more likely to be there when the intercepts occurred, hence inflating the sample average number of trips relative to the population's average. Third, the observed trips (OB) are slightly higher (12.32) than the number of trips anticipated by the survey respondents for 2001, suggesting relatively stable demand for visits to Clear Lake between 2000 and 2001. Fourth, and finally, the anticipated number of trips for 2001 decrease, as expected, with the total cost per trip, from an average number of trips just under 12 per year under current conditions (CB_0) to approximately 7.5 trips per year given an average cost increase of $17 per trip ($CB_2$). Thus, households appear to be responding to the hypothetical price increase at least in the direction expected.

Table 1. Summary Statistics

Variable	Mean	Std. Dev.	Minimum	Maximum
OB trips (y_{i1})	12.32	11.86	1	52
CB_0 trips (y_{i2})	11.71	11.77	0	50
CB_1 trips (y_{i3})	10.29	10.86	0	50
CB_2 trips (y_{i4})	7.33	9.12	0	50
Travel Cost ($P_{i1} = P_{i2}$)	$56.73	$56.62	$5.37	$512.50
Travel Cost + $B ($P_{i3}$)	$64.00	$57.54	$8.37	$522.50
Travel Cost + $C ($P_{i4}$)	$73.61	$58.90	$12.37	$537.50
Household Income (I_i)	$59,752	$37,713	$7,500	$200,000
Male	0.63	0.48	0	1
Age	43.62	13.59	15	82
Education	0.74	0.44	0	1
Number of Household Members	3.07	1.40	1	9

Turning to the socio-demographic data, we find that the percentage of males (63%), average household income, and level of education are higher in the sample than in the Iowa population as a whole. This, in part, is also a consequence of the on-site nature of the survey process, as frequent recreationists are more likely to be included in the sample and these, in turn, are more likely to be males with a higher level of income and education.

In estimating the MPLN model using the Clear Lake data, several restrictions were imposed on the form of the λ_{ij}'s (i.e. expected trips). First, we assume that the β's in equation (28) are the same across the three contingent behavior trips, with expected trips changing only due to changes in the corresponding price levels. Second, we assume that the socio-demographic factors (other than income) impact the expected number of trips in the same way for both observed trips and the three contingent trips.[10] The resulting functional forms for the λ_{ij}'s are given by:

$$\lambda_{ij} = \begin{cases} \exp\left(\beta_{0,OB} + \beta_{P,OB} P_{i1} + \beta_{I,OB} I_i + \delta' z_i\right) & j = 1 \\ \exp\left(\beta_{0,CB} + \beta_{P,CB} P_{ij} + \beta_{I,CB} I_i + \delta' z_i\right) & j = 2,3,4. \end{cases} \quad (29)$$

Finally, we also impose a restriction on the structure of the variance-covariance matrix for the MPLN model. Specifically, we assume that Ω in equation (14) is given by

$$\begin{aligned} \Omega &= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ & & \sigma_3^2 & \sigma_{34} \\ & & & \sigma_4^2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_O^2 & \rho_{OC}\sigma_O\sigma_C & \rho_{OC}\sigma_O\sigma_C & \rho_{OC}\sigma_O\sigma_C \\ & \sigma_C^2 & \rho_{CC}\sigma_C^2 & \rho_{CC}\sigma_C^2 \\ & & \sigma_C^2 & \rho_{CC}\sigma_C^2 \\ & & & \sigma_C^2 \end{bmatrix}. \end{aligned} \quad (30)$$

---

[10] A more general specification allowing the demographic effects to differ between observed trips and contingent trips was estimated, but the differences between the OB and CB parameters were not statistically different as a group based on a likelihood ratio test.

This implies that the unobserved error component for the three contingent trips (CB_0, CB_1, and CB_2) have the same covariances with each other and with the observed trip data.

IV. Results

Table 2 provides the estimates of the MPLN model.[11] We present estimates both with and without the correction for on-site sampling. Several patterns emerge in the results. First, the price and income coefficients have the expected signs and are statistically significant at a one percent level for both observed and contingent behavior trips. All else equal, an increase in travel cost decreases the expected number of trips, whereas trips increase with income. Second, these coefficients (i.e., the $\beta's$) do not differ substantially between the observed and contingent trips. However, the price responsiveness is lower among the contingent trips than for the observed trips, whereas contingent trips are more sensitive to income than observed trips. Third, the price and income coefficients do not change substantially with the correction for on-site sampling, though they are generally smaller in size.

Turning to the socio-demographic characteristics, the results are less consistent across the corrected and uncorrected models. For the MPLN specification corrected for on-site sampling, all of the socio-demographic characteristics (except the number of household members) are statistically significant and have the expected signs. Men are found to take significantly more recreational trips to Clear Lake than women and the relationship between age and trips is quadratic, with the young and old taking more trips than middle aged

[11] The MPLN model was estimated using maximum simulated likelihood following Munkin and Trivedi (1999). Hess, Train and Polak (2003) develop a new simulation technique using randomly shifted and shuffled uniform vectors. We employ this technique using 1000 draws in the simulation. The authors would like to thank Kenneth Train for suggesting this method of simulation and also thank Stephane Hess for providing the gauss code and suggestions for implementation.

individuals. Having attended college decreases recreational trips. For the uncorrected specifications, the socio-demographic coefficients are generally less significant.

Table 2. Multivariate Poisson LogNormal Models
(Standard Errors in Parentheses)[a]

Parameter	Corrected for On-Site Sampling	Not Corrected for On-Site Sampling
$\beta_{0,OB}$	0.74** (0.09)	1.94** (0.04)
$\beta_{0,CB}$	0.55** (0.09)	1.57** (0.05)
$\beta_{P,OB}$	-1.57** (0.07)	-1.58** (0.07)
$\beta_{P,CB}$	-1.48** (0.05)	-1.66** (0.06)
$\beta_{I,OB}$	0.95** (0.10)	1.08** (0.09)
$\beta_{I,CB}$	1.06** (0.08)	1.31** (0.08)
Male	27.40** (4.45)	9.71 (5.05)
Age	-4.20** (0.69)	-2.98** (0.94)
Age^2	0.04** (0.007)	0.03** (0.01)
School	17.82** (4.45)	13.04* (6.28)
Household	-4.03* (1.77)	0.77 (2.40)
σ_O	1.17** (0.04)	0.95** (0.03)
σ_C	1.26** (0.04)	1.10** (0.03)
ρ_{OC}	0.95** (0.006)	0.92** (0.01)
ρ_{CC}	0.99** (0.002)	0.98** (0.004)
LogLik	-6,153.39	-6,105.32

*Significant at 5% level; **significant at 1% level.
[a]All of the parameters are scaled by 100, except the constants (which are unscaled), and the income coefficient (which is scaled by 100,000).

Finally, it is worth noting the parameters associated with the mixing distribution. For the MPLN model, we clearly reject both equidispersion and independence of the observed and contingent trip data. The correlation among the trips is high, with both ρ_{OC} and ρ_{CC} estimated to be positive and close to one. Both σ_O and σ_C are significantly different from zero, indicating overdispersion in the data.

The parameter estimates in Table 2 can be used to illustrate implications of the models in terms of trip behavior and the implied welfare gains associated with each trip. Table 3a provides estimates of the consumer surplus per trip calculated as $CS_j = \beta_{P,j}^{-1}$ for both observed trips (j=1) and predicted trips for 2001 (j=2). Both models the corrected and uncorrected predict roughly the same consumer surplus per trip, ranging from \$60 to \$68. Correcting for the on-site sampling leads to a somewhat larger surplus measure, with an increase of 12% for predicted trips.

The big impact, however, from correcting for on-site sampling comes in the form of the predicted number of trips. Table 3b provides estimates of the population average trips. For the MPLN model this corresponds to δ_{ij} in equation (16). As expected, there is a substantial difference between the average numbers of trips when the model is corrected for on-site sampling versus when it is not. Without this correction, average trips range from 13.43 to 14.24. This is consistent with the sample averages reported in Table 1. However, correcting for the on-site sampling, we see a substantial drop in the estimated average number of trips in the population. For the MPLN model the average is reduced by two-thirds to only five trips per household. The estimates in Table 3b are based upon the average household characteristics (i.e., age, income, education, etc.) found in the survey sample.

However, these too are biased by the on-site sampling process. Table 3c recalculates the estimated average number of trips using population averages for the explanatory variables drawn from the 2000 census data for Iowa households. The average number of trips per household drops further as a result to under one and a half trips per household.

Table 3. Fitted Trips and Consumer Surplus Measures from the MPLN Model

	Corrected for On-Site Sampling	Not Corrected for On-Site Sampling
	a. Consumer Surplus Per Trip	
CS_1	63.72** (2.77)	63.47** (2.77)
CS_2	67.80** (2.51)	60.42** (2.14)
	b. Fitted Population Trips	
$E[y_{i1} \mid x_{i1}]$	5.51 (11.15)	14.24 (20.42)
$E[y_{i2} \mid x_{i2}]$	5.63 (13.14)	13.43 (25.11)
	c. Fitted Population Trips (corrected for population characteristics)	
$E\left[y_{i1} \mid x_{i1}^P\right]$	1.28 (2.45)	
$E\left[y_{i2} \mid x_{i2}^P\right]$	1.39 (2.99)	

Finally, there are a number of hypothesis tests of interest. The first of the hypothesis tests we consider constrains the parameters of the observed and contingent behavior trip functions to be the same; i.e., $\beta_{k,O} = \beta_{k,C}$, $k = 0, P, I$. The results are reported in column three of Table 4. In general, the resulting parameters are a compromise between the observed and contingent behavior parameters, but the hypothesis itself is clearly rejected with a p-value of less than 0.001.

Table 4. Hypothesis Tests Using Multivariate Poisson-Lognormal Model
(Standard Errors in Parentheses)[a]

Parameter	Unrestricted	Consistency $\beta_{k,O}=\beta_{k,C},\ k=0,P,I$	Restricted Correlation
$\beta_{0,OB}$	0.74^{**}		0.29^{*}
	(0.09)	0.68^{**}	(0.12)
$\beta_{0,CB}$	0.55^{**}	(0.07)	0.16
	(0.09)		(0.12)
$\beta_{P,OB}$	-1.57^{**}		-1.67^{**}
	(0.07)	-1.46^{**}	(0.08)
$\beta_{P,CB}$	-1.48^{**}	(0.05)	-1.50^{**}
	(0.05)		(0.07)
$\beta_{I,OB}$	0.95^{**}		1.30^{**}
	(0.10)	0.99^{**}	(0.15)
$\beta_{I,CB}$	1.06^{**}	(0.08)	1.40^{**}
	(0.08)		(0.15)
Male	27.40^{**}	26.72^{**}	17.71
	(4.45)	(5.03)	(10.81)
Age	-4.20^{**}	-4.58^{**}	-5.87^{**}
	(0.69)	(0.91)	(1.27)
Age^2	0.04^{**}	0.04^{**}	0.06^{**}
	(0.007)	(0.009)	(0.01)
School	17.82^{**}	18.07^{**}	15.73
	(4.45)	(4.86)	(9.25)
Household	-4.03^{*}	-5.17^{**}	1.15
	(1.77)	(1.82)	(3.47)
σ_O	1.17^{**}	1.20^{**}	
	(0.04)	(0.03)	1.25^{**}
σ_C	1.26^{**}	1.23^{**}	(0.04)
	(0.04)	(0.03)	
ρ_{OC}	0.95^{**}	0.94^{**}	
	(0.006)	(0.006)	
ρ_{CC}	0.99^{**}	0.99^{**}	
	(0.002)	(0.003)	
LogLik	-6,153.39	-6,237.72	-6,259.33
$\chi^2_{df=3}$		168.66	

*Significant at 5% level; **significant at 1% level.

[a]All of the parameters are scaled by 100, except the constants (which are unscaled), and the income coefficient (which is scaled by 100,000).

The second restricted version of the model replaces the multivariate lognormal mixing distribution with a single lognormal variable (i.e., $\varepsilon_{ij} = \varepsilon_i \sim N(0,\sigma^2)\forall j$). Essentially, we are restricting $\sigma_O = \sigma_C$ and $\rho_{OC} = \rho_{CC} = 1$. This mimics the structure of the MPG distribution, but uses a lognormal mixing distribution rather than a gamma one. While this model represents a boundary restriction on correlation parameters, making a standard likelihood ratio test problematic, the large reduction in the log-likelihood function suggests little support for this alternative specification.

V. Conclusions

On-site samples are frequently used in recreation demand analysis to insure that users of the site in question are represented in the sample. It has long been recognized that this results in a sample that is both truncated and endogenously stratified with respect to the respondents' reported trips to the site. The correction procedures that have been previously developed focused on observed trip data alone (e.g., Shaw, 1988, and Englin and Shonkwiler, 1995). However, researchers are frequently incorporating contingent behavior questions into their recreation demand surveys as well, asking households to indicate their future trip plans and how their trips might change given price or quality changes to the site in question (See, e.g., Rosenberger and Loomis, 1999; Azevedo, Herriges, and Kling, 2003; and Grijalva, *et al.* 2002). While the contingent behavior trip responses are not directly truncated or endogenously stratified, they are impacted indirectly through their correlation with observed trips. The contingent behavior data, like its observed counterpart, will not be representative of the population as a whole. In this paper, we have presented an extension of Shaw's (1988) correction to a multivariate setting using the MPLN model.

The empirical analysis, using data from an intercept survey at Clear Lake in northcentral Iowa, indicates that the failure to correct for on-site sampling procedures results in substantial bias in the estimated average number of trips to the site, both observed and contingent, overstating population trip levels by a factor of 11. The impact on the estimated consumer surplus per trip is somewhat small. We also reject the hypothesis that the observed and contingent trips follow exactly the same demand structure, but the differences, while statistically significant, appear to be minor.

VI. References

Aitchison, J., and C.H. Ho (1989). "The Multivariate Poisson-Log Normal Distribution." *Biometrika* 76: 643-53.

Arbous, A.G., and J.E. Kerrich (1951). "Accident statistics and the concept of accident proneness. *Biometrika* 37:340-431.

Azevedo, C.D., J.A. Herriges, and C.L. Kling (2003). "Combining Revealed and Stated Preferences: Consistency Test and Their Interpretations." *American Journal of Agricultural Economics* 85(August): 525-37.

Bates, G.E., and J. Neyman (1952). "Contributions to the theory of accident proneness I: An optimistic model of the correlation between light and severe accidents." *University of California Publications in Statistics* 1:212-53.

Creel, M.D., and J.B. Loomis. 1990. "Theoretical and Empirical Advantages of Truncated Count Data Estimators for Analysis of Deer Hunting in California." *American Journal of Agricultural Economics* 72 (May): 434-41.

Englin, J., and T.A. Cameron (1996). "Augmenting Travel Cost Models with Contingent Behavior Data; Poisson Regression Analyses with Individual Panel Data." *Environmental and Resource Economics* 7:133-47.

Englin, J., J.B. Loomis, and A. Gonzalez-Caban. 2001. "The dynamic path of recreational values following a forest fire: a comparative analysis of states in the Intermountain West." *Canadian Journal of Forest Research* 31(10):1837-44.

Englin, J., and J.S. Shonkwiler (1995). "Estimating Social Welfare Using Count Data Models: An Application to Long-Run Recreation Demand under Conditions of Endogenous Stratification and Truncation." *The Review of Economics and Statistics* 77(1): 104-12.

Grijalva, T.C., R.P. Berrens, A.K. Bohara, and W.D. Shaw (2002). "Testing the Validity of Contingent Behavior Trip Responses." *American Journal of Agricultural Economics* 84(2): 401-14.

Grogger, J.T., and R.T. Carson. 1991. "Models for Truncated Counts." *Journal of Applied Econometrics* 6(3) (September): 225-238.

Hausman, J., B.H. Hall, and Z. Griliches (1984). "Econometric Models for Count Data with an Application to the Patents-R & D Relationship." *Econometrica* 52(4): 909-38.

Hess, S., K. Train, and J. Polak (2003). "On the use of randomly shifted and shuffled uniform vectors in the estimation of a Mixed Logit model for vehicle choice." Working paper.

Laitila, T. (1999). "Estimation of Combined Site-Choice and Trip-Frequency Models of Recreational Demand using Choice-Based and On-Site Samples" *Economics Letters* 64: 17-23.

Munkin, M.K., and P.K. Trivedi (1999). "Simulated Maximum Likelihood Estimation of Multivariate Mixed-Poisson Regression Models, with Application." *Econometrics Journal* 2: 29-48.

Nelson, J.F. (1985). "Multivariate Gamma-Poisson Models." *Journal of the American Statistical Association* 80:254-83.

Rosenberger, R.S., and J.B. Loomis. (1999). "The Value of Ranch Open Space to Tourists: Combining Observed and Contingent Behavior Data." *Growth and Change* 30: 366-83.

Shaw, D. (1988). "On-site Samples' Regression: Problems of Non-Negative Integers, Truncation, and Endogenous Stratification." *Journal of Econometrics* 37: 211-23.

Shonkwiler, J.S. (1995). "Systems of Travel Cost Models of Recreation Demand." W-133 paper.

Smith, V.K., and W.H. Desvousges (1985). "The Generalized Travel Cost Model and Water Quality Benefits: A Reconsideration" *Southern Economic Journal* 52:371-81.

Whitehead, J.C., T.C. Haab, and J. Huang. (2000). "Measuring recreation benefits of quality improvements with revealed and stated behavior data." *Resource and Energy Economics* 22: 339-54.

Winkelmann, Rainer (2000). *Econometric Analysis of Count Data*. (Springer-Verlag: New York).

CHAPTER 3. COMBINING REVEALED PREFERENCE AND TWO STATED PREFERENCE DATA: CONTINGENT BEHAVIOR AND CONTINGENT VALUATION

A paper to be submitted to a journal in the field

Kevin J. Egan

I. Introduction

The travel cost model was one of the first models used to measure non-market environmental goods. The model is applied extensively but used alone, with only revealed preference data, it is difficult to measure a quality change. Quality changes may take decades to materialize. Policymakers would like welfare benefit information before the proposed quality improvements are undertaken. Thus, stated preference information such as contingent valuation and contingent behavior questions were key approaches to gain this valuable information ex-ante.

More recently researchers asked both travel cost and contingent valuation questions to take advantage of the strengths of each type of data. Cameron (1992) was the first paper to combine the two sources of information into one joint model. There are many reasons to combine travel cost and contingent valuation data. Maybe the most compelling reason is the increased precision garnered from using more information to estimate the parameters. The revealed preference information imposes the discipline of the market on the stated preference data while allowing stated preference data to fill-in some information about preferences not captured by revealed preference data.

Also, the same individuals are answering both types of questions. It is theoretically possible to model all the data as being derived from one set of preferences. However

Cameron's paper is not a utility theoretic model. She used an ad-hoc error structure since both her utility difference function for the contingent valuation data and her demand function for the travel cost data had additive errors, impossible to derive from one another. Huang, Haab, and Whitehead (1997) proposed a model that is utility theoretic. This model is discussed in detail in the next section.

Layman, Boyce, and Criddle (1996) was the first paper to measure a quality change using contingent behavior questions as an alternative to contingent valuation. The authors argued many advantages of using contingent behavior questions. One is that contingent behavior data is identical in form to travel cost data, therefore, hypothetical trips are easier for the visitor to understand as an "ordinary commodity" with a price (travel cost) and substitutes (other lakes).

Huang, Haab, and Whitehead (1997) combine all three types of information (RP and the two SP data, contingent behavior and contingent valuation) for a quality change. This paper will utilize Huang, Haab, and Whitehead's model to analyze information from a portion of the Clear Lake survey focusing on visitors' responses to water quality improvement scenarios.

Another modeling issue arises due to the sample being collected on-site. The observed behavior data (actual reported trips) are truncated at one (excluding non-users) and endogenously stratified (over sampling those individuals who are more frequent users of the site). The contingent behavior data (anticipated trips given the improved quality conditions) are incidentally impacted as the over-sampled individuals who took higher actual trips are also more likely to anticipate taking higher contingent behavior trips. Therefore, the contingent behavior data is incidentally over-sampled.

This paper also utilizes contingent valuation data which again is incidentally impacted since the individuals answering the contingent valuation questions are not representative of the population. To get unbiased estimates one needs a joint model that corrects for the incidental truncation and endogenous stratification. In this paper I extend Huang, Haab, and Whitehead's joint model by adapting the correction for on-site sampling derived by Shaw (1988). Due to the on-site sampling, then the joint model not only utilizes more information but it is a necessity to get unbiased estimates from the contingent valuation data.

Section II derives the analytical model and section III details the estimation procedures with the travel cost model estimated first to compute annual consumer surplus. Then the bid function approach will be employed to estimate WTP from the contingent valuation questions. Finally a theoretically consistent model will be derived that jointly utilizes the RP and the two SP information, contingent behavior and contingent valuation data.

II. Analytical Model for Combining Contingent Behavior and Contingent Valuation Data

This section's discussion follows closely the joint model in Huang, Haab, and Whitehead (1997) [HHW]. To begin, define the visitor's willingness to pay for the quality improvement as an equivalent variation measure:

$$WTP = e\left(p,q,u\right) - e\left(p,q^{*},u\right) \tag{1}$$

where $e(\cdot)$ is the expenditure function, p is the price of a recreation trip (the travel cost), m is income, q is the current level of quality, and q^{*} is the improved level of quality. The reference level of utility is $u = \left(v\left(p, q^*, m\right)\right)$ implying the visitor's property rights are with the future improved quality level. In the survey it is described to the visitor as a future improvement, but it could easily be depicted as a return to some historical level of water quality since the lake has been deteriorating for 50 years. Therefore, equivalent variation is the appropriate measure by establishing some historically higher quality level as the reference point.

Substituting $u = v\left(p, q^*, m\right)$ into the WTP variation function equation (1) yields:

$$w = e\left[p, q, v\left(p, q^*, m\right)\right] - m. \tag{2}$$

Assuming that q is a normal good, the partial derivative of equation (2) with respect to income is then:

$$\frac{\partial w}{\partial m} = \frac{\partial e(q)}{\partial v} \frac{\partial v\left(q^*\right)}{\partial m} - 1 = \mu - 1 > 0 \tag{3}$$

where e_v is the marginal cost of utility evaluated at q, v_m^* is the marginal utility of income evaluated at q^*, and $\mu = e_v v_m^* > 1$. The marginal utility of income transfers dollars into "utils" at the margin at the higher quality level and the marginal cost of utility transfers "utils" back into dollars at the margin but at the degraded quality level. When evaluated at the same quality level $v_m^* = 1/e_v^*$ and the transfer of "utils" to dollars and vice-versa is equivalent. Now μ can be defined as $\mu = e_v / e_v^*$. When quality is a normal good, then the marginal cost of utility is greater with the degraded quality and the income effect will be positive. If the marginal utility of income is constant then μ equals one and the income effect is zero[1].

The partial derivative of equation (2) with respect to p is:

$$\frac{\partial w}{\partial p} = \frac{\partial e(q)}{\partial p} + \frac{\partial e(q)}{\partial v}\frac{\partial v(q^*)}{\partial p} = e_p + e_v v_p^* < 0. \tag{4}$$

Using Shephard's Lemma and Roy's Identity respectively:

$$\frac{\partial e(p,u)}{\partial p} = e_p = x^h$$

$$-\frac{v_p^*}{v_m^*} = x^{m*} \Rightarrow v_p^* = -v_m^* x^{m*}$$

and substituting them into equation (4) yields:

$$\begin{aligned} \frac{\partial w}{\partial p} &= x^h - e_v v_m^* x^{m*} \\ &= x^h - \mu x^{m*} \end{aligned} \tag{5}$$

where x^h is the hicksian demand at the current quality level and x^{m*} is the marshallian demand at the higher quality level. Since at the original level of quality, $x^h = x^m$, then:

$$\frac{\partial w}{\partial p} = x^m - \mu x^{m*}. \tag{6}$$

The partial derivatives (equations (3) and (6)) can be used to derive the link between the visitor's contingent valuation and contingent behavior responses. Assume a linear WTP function as HHW have done:

$$w_i = \alpha + \beta p_i + \lambda m_i + \sigma_1 \varepsilon_{1i} \tag{7}$$

---

[1] See Whitehead (1995, p. 209).

where ε_{1i} is the normally distributed error term. Now take the partial derivative of equation (7) with respect to own price and income and then set the results equal to the previous derivations (i.e. equations (3) and (6)):

$$\frac{\partial w}{\partial p} = \beta = x^m - \mu x^{m*} \qquad (8)$$

$$\frac{\partial w}{\partial m} = \lambda = \mu - 1. \qquad (9)$$

Substituting μ-1 for λ in equation (7) yields:

$$w_i = \alpha + \beta p_i + (\mu - 1) m_i + \sigma_1 \varepsilon_{1i}. \qquad (10)$$

Solving equation (8) for x^m:

$$x^m = \beta + \mu x^{m*}$$

and letting x_i denote the visitor's revealed number of trips to Clear Lake over the last year and $x_i + \sigma_2 \varepsilon_{2i}$ denote the visitor's stated number of trips under the improved water quality scenario (where ε_{2i} is the visitor's measurement error from the mean stated number of trips (x^*)), then equation (8) can be written as:

$$x_i = \beta + \mu\left(x_i^* + \sigma_2 \varepsilon_{2i}\right). \qquad (11)$$

The previous two equations:

WTP Variation Function: $w_i = \alpha + \beta p_i + (\mu - 1) m_i + \sigma_1 \varepsilon_{1i}$ (10)

Trip Change Function: $x_i = \beta + \mu\left(x_i^* + \sigma_2 \varepsilon_{2i}\right)$ (11)

are theoretically derived functions which will be used to measure the quality change. As HHW note, "stated and revealed preference for a quality improvement are analytically consistent since the own-price effects on WTP are directly related to the measure of recreation behavior change."

To gain more intuition about the functions consider the special case in which μ equals one, (i.e. meaning the marginal utility of income is constant). The above functions reduce to:

$$\text{WTP Variation Function: } w_i = \alpha + \beta p_i + \sigma_1 \varepsilon_{1i}$$
$$\text{Trip Change Function: } x_i = \beta + x_i^* + \sigma_2 \varepsilon_{2i}.$$

Solving for the expected additional trips taken in response to the improvement in environmental quality provides a convenient interpretation of β:

$$\begin{aligned} E\left(x_i - x_i^*\right) &= E\left(\beta + \sigma_2 \varepsilon_{2i}\right) \\ &= \beta \end{aligned}$$

i.e., β equals the expected additional trips taken in response to the improvement in environmental quality.

III. Estimation Procedures

This section begins by discussing estimation procedures for the travel cost model, which jointly models the observed behavior and contingent behavior data as well as correcting for on-site sampling. Next to be discussed is the estimation procedure for the WTP variation function separately, and then the joint estimation procedure of the trip change function and the WTP variation function corrected for on-site sampling.

A. Travel Cost Model

Assume the observed (x_i) and contingent (x_i^*) trips from the on-site sample are conditionally (given the independent variables z_i and z_i^*) bivariate Poisson-lognormally distributed[2]

$$f_{OS}\left(x_i, x_i^* \mid z_i, z_i^*\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{x_i}{\lambda_i \exp\left(\frac{1}{2}\sigma^2\right)} \frac{\exp\left(\lambda_i e^{\varepsilon_i}\right)\left(\lambda_i e^{\varepsilon_i}\right)^{x_i}}{x_i!} \frac{\exp\left(\lambda_i^* e^{\varepsilon_i^*}\right)\left(\lambda_i^* e^{\varepsilon_i^*}\right)^{x_i^*}}{x_i^*!} \times \frac{1}{2\pi\sigma\sigma^*\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2\left(1-\rho^2\right)}\left\{\left(\frac{\varepsilon_i}{\sigma}\right)^2 - 2\rho\frac{\varepsilon_i\varepsilon_i^*}{\sigma\sigma^*} + \left(\frac{\varepsilon_i^*}{\sigma^*}\right)^2\right\}\right] d\varepsilon_i d\varepsilon_i^* \tag{12}$$

where the expected trips are specified as

$$\begin{aligned} \lambda_i &= \exp(\alpha + \beta p_i + \gamma m_i + \sigma\varepsilon_i) \\ \lambda_i^* &= \exp(\alpha^* + \delta^* D + \beta^* p_i + \gamma^* m_i + \sigma^* \varepsilon_i) \end{aligned} \tag{13}$$

allowing different coefficients for the observed and contingent trips. Specifically, x_i is the total number of recreation trips to Clear Lake from November 1999 to October 2000 and x_i^* is the total number of recreation trips to Clear Lake reported under plan B, a proposed water quality improvement. A dummy variable *(D)* is included for the contingent behavior data since the visitors were sent two different water quality improvement scenarios, one describing a moderate water quality improvement $(D = 0)$ and the other describing a larger improvement $(D = 1)$. The visitor's income is m_i , and p_i is the price of a recreation trip. It is estimated by the equation:

$$p_i = c_i + \eta T_i W_i \tag{14}$$

---

[2] See Chapter 2 of this dissertation for derivation of the multivariate Poisson-lognormal model corrected for on-site sampling.

where c_i is the visitor's out-of-pocket travel cost. The visitor's round trip travel time is T_i, and W_i denotes the wage rate. The proportion at which the travel time is valued is represented by η. For simplicity assume η is a fixed fraction equal to one-third[3]. In calculating p_i in this way it is assumed the visitors are able to choose hours worked at the margin. Another simplifying assumption is that all trips to the lake are for roughly the same amount of time, or that length of stay at the lake is inconsequential in the modeling process.

B. WTP Variation Function

The bid function approach (Cameron, 1988) will be used to estimate the WTP variation function (equation (10)). The visitor can be expected to answer yes to the referendum format contingent valuation question if her true willingness to pay, w_i, is more than the bid value, B_i. Thus, the probability the visitor will say yes is:

$$\begin{aligned}
\text{Pr(yes)} &= \Pr\left(w_i > B_i\right) = \Pr\left(\alpha + \beta p_i + (\mu - 1) m_i + \delta_1 D + \sigma_1 \varepsilon_{1i} > B_i\right) \\
&= \Pr\left(\varepsilon_{1i} > \frac{B_i - \alpha - \beta p_i - (\mu - 1) m_i + \delta_1 D}{\sigma_1}\right) \\
&= 1 - \Phi\left(\frac{B_i - \alpha - \beta p_i - (\mu - 1) m_i + \delta_1 D}{\sigma_1}\right) \\
&= \Phi\left(\frac{-B_i + \alpha + \beta p_i + (\mu - 1) m_i + \delta_1 D}{\sigma_1}\right)
\end{aligned}$$

where Φ denotes the standard normal cdf, and again added is D, the dummy variable representing the two versions of the survey. The probability the visitor will say no is simply the complement to the above probability. Letting I_i be the indicator variable which equals

---

[3] Cesario (1976) suggested valuing travel time at one-third the wage rate.

one if the survey respondent answers yes and equals zero otherwise, then the log-likelihood function can be written as:

$$LL = \sum_{i=1}^{n} \log\left\{\Phi\left[q_i\left(\frac{-B_i + \alpha + \beta p_i + (\mu - 1)m_i + \delta_1 D}{\sigma_1}\right)\right]\right\} \tag{15}$$

where $q_i = 2I_i - 1$.

By the method of maximum likelihood estimation the coefficients $\tilde{\alpha}, \tilde{\beta}, \tilde{\mu}, \tilde{\delta}_1, \tilde{\sigma}_1$ are estimated.[4] Substituting the coefficients back into the WTP variation function:

$$\begin{aligned} \tilde{w}_i^{CV} &= \tilde{\alpha} + \tilde{\beta} p_i + (\tilde{\mu} - 1) m_i + \tilde{\delta}_1 D \\ \approx{w}^{CV} &= \frac{\sum_{i=1}^{n} \tilde{w}_i^{CV}}{n} \end{aligned} \tag{16}$$

where $\overset{\simeq}{w}^{CV}$ is one of three estimates this paper will be calculating of the visitor's willingness to pay for the quality improvement. This WTP estimate is labeled CV since it uses the contingent valuation question. Note however, unlike the usual bid function approach, the WTP variation function does not exclusively use contingent valuation data since p_i, the travel cost, is included as an explanatory variable.

However, when the sample is collected on-site the WTP estimates from the bid function approach may be biased due to the sample not being representative of the population. If visitors with high observed trips (i.e. those who are over-sampled) are more likely to answer yes to the contingent valuation question, then the WTP estimates from the bid function approach will be biased upward. One way to correct the contingent valuation

---

[4] A quadratic price coefficient was estimated, however it was not significant, and the hypothesis of excluding the term could not be rejected based on a likelihood ratio test.

estimates is to estimate a joint model like the one HHW have derived, where the joint density is corrected for on-site sampling.

C. Joint Estimation

To begin, I will discuss the joint estimation assuming a random population sample and then I will discuss the correction to the log-likelihood function for on-site sampling. To jointly model equations (10) and (11), both functions will be combined into one log-likelihood function where the correlated errors will be accounted for by assuming a bivariate normal distribution, $(\varepsilon_{1i}, \varepsilon_{2i}) \sim N\left(0, 0, \sigma_1^2, \sigma_2^2, \rho\right)$. Then if the parameters are restricted to be equal, the log-likelihood function will be estimated resulting in one willingness to pay estimate.

The joint distribution for visitor i is then:

$$\begin{bmatrix} w_i \\ x_i \end{bmatrix} \sim N\left(\begin{bmatrix} \alpha + \beta p_i + (\mu - 1) m_i + \delta_1 D \\ \beta + \mu x_i^* + \mu \delta_2 D \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \mu\sigma_{12} \\ \mu\sigma_{12} & \mu^2\sigma_2^2 \end{bmatrix}\right)$$

where $\sigma_{12} = \dfrac{\rho}{(\sigma_1\sigma_2)}$. Following HHW the joint distribution can be written as the distribution of w_i conditional on x_i multiplied by the distribution of x_i: $f\left(w_i, x_i\right) = f\left(w_i \middle| x_i\right) f\left(x_i\right)$.

The conditional distribution of w_i is:

$$w_i \Big| x_i \sim N\left(\alpha + \beta p_i + (\mu - 1) m_i + \delta_1 D + \rho\sigma_1\left(\frac{x_i - \beta - \mu x_i^* + \mu\delta_2 D}{\mu\sigma_2}\right), \left(1 - \rho^2\right)\sigma_1^2\right).$$

The joint distribution combines the continuous trips with the discrete WTP responses. This distribution can be written as the product of a Bernoulli distribution conditional on the trip decision and the density function of trips:

$$P(yes|x_i)^{I_i} P(no|x_i)^{1-I_i} f(x_i) = f(x_i) P(w_i \geq T_i|x_i)^{I_i} P(w_i < T_i|x_i)^{1-I_i}$$

$$= \phi\left(\frac{x_i - \beta - \mu x_i^* - \mu\delta_2 D}{\mu\sigma_2}\right)\Phi\left(\frac{(-B_i + \alpha + \beta p_i + (\mu-1)m_i + \delta_1 D)/\sigma_1 + \rho(x_i - \beta - \mu x_i^* - \mu\delta_2 D)/\mu\sigma_2}{(1-\rho^2)^{\frac{1}{2}}}\right)^{I_i}$$

$$\left[1 - \Phi\left(\frac{(-B_i + \alpha + \beta p_i + (\mu-1)m_i + \delta_1 D)/\sigma_1 + \rho(x_i - \beta - \mu x_i^* - \mu\delta_2 D)/\mu\sigma_2}{(1-\rho^2)^{\frac{1}{2}}}\right)\right]^{1-I_i}$$

where I_i equals one if the survey respondent indicated yes to the contingent valuation question and equals zero otherwise, and ϕ and Φ are the normal density and cumulative distribution functions. The log-likelihood function is then:

$$LL = -n\ln\left(\sigma_2\sqrt{2\pi}\right) - \frac{1}{2\sigma_2^2}\sum_{i=1}^{n}\left(\frac{x_i}{\mu} - \frac{\beta}{\mu} - x_i^* - \delta_2 D\right)^2$$
$$+\sum_{i=1}^{n}\ln\left\{\Phi\left[q_i\left(\frac{(-B_i + \alpha + \beta p_i + (\mu-1)m_i + \delta_1 D)/\sigma_1 + \rho\left(\frac{x_i}{\mu} - \frac{\beta}{\mu} - x_i^* - \delta_2 D\right)/\sigma_2}{(1-\rho^2)^{\frac{1}{2}}}\right)\right]\right\} \quad (17)$$

where $q_i = 2I_i - 1$. The first line is the log-likelihood function for estimating the trip change function. The second line is the log-likelihood function for estimating the conditional distribution of the WTP variation function (i.e. $w_i|x_i$).

Shaw (1988) corrects for truncation and endogenous stratification by calculating the on-site sample's density function. Shaw assumes that visitors taking x_i trips are x_i times more likely to be intercepted than someone who takes only one trip. Using this assumption, he shows the on-site sample's density function can be written as

$$f_{OS}(x_i \mid z_i) = \frac{x_i}{E[x_i \mid z_i]} f(x_i \mid z_i), \quad (18)$$

where the population density is reweighted by the ratio of the observed value and the expected value. If the distribution is normal then

$$E[x_i \mid z_i] = \theta' z_i + \sigma Q(d_i) \tag{19}$$

where $d_i = \theta' z_i / \sigma$, $r_i = \phi(d_i)/\Phi(d_i)$, $Q(d_i) = 1/(d_i + r_i)$, $\phi(d_i)$ and $\Phi(d_i)$ are, respectively the standard normal density and cumulative distribution function evaluated at d_i. The reweighting procedure Shaw derives in equation (18) and (19) can be applied to HHW's trip change function such that the joint log-likelihood function is

$$\begin{aligned} LL = {} & -n\ln\left(\sigma_2^2\sqrt{2\pi}\right) - \frac{1}{2\sigma_2^2}\sum_{i=1}^{n}\left(\frac{x_i}{\mu} - \frac{\beta}{\mu} - x_i^* - \delta_2 D\right)^2 \\ & + \sum_{i=1}^{n}\ln x_i - \sum_{i=1}^{n}\ln\left[d_i\Phi(d_i) + \phi(d_i)\right] \\ & + \sum_{i=1}^{n}\ln\left\{\Phi\left[q_i\left(\frac{\left(-B_i + \alpha + \beta p_i + (\mu - 1)m_i + \delta_1 D\right)/\sigma_1 + \rho\left(\frac{x_i}{\mu} - \frac{\beta}{\mu} - x_i^* - \delta_2 D\right)/\sigma_2}{\left(1-\rho^2\right)^{\frac{1}{2}}}\right)\right]\right\} \end{aligned} \tag{20}$$

IV. The Data

In the summer of 2000 visitors to Clear Lake were intercepted at the boat ramps, beaches, and fishing docks. A total of 1,024 intercepted visitors agreed to participate in the mail survey which occurred in October, 2000. The visitors were paid \$5 for a returned survey. Of the deliverable surveys 626 were returned resulting in a 62.7% response rate.

The survey was conducted to measure visitor's and local resident's willingness to pay for quality improvements to Clear Lake. The visitor's survey contains different quality improvement plans, and this paper focuses on one in particular, Plan B, which consisted of a moderate and a high water quality improvement. However all of the analysis is easily extended to the other plans. See Azevedo, Herriges, Kling (2001) for survey summary statistics.

There are three questions of interest in the survey.[5] The visitors were first asked a revealed preference question. It asked the visitors to report the number of trips they had taken to Clear Lake over the last year. They were then asked two stated preference questions. The first being a contingent behavior question asking them for the number of trips they would have taken over the past year to Clear Lake if conditions were as described under Plan B, the proposed water quality improvement. Second, the visitors were asked a referendum format contingent valuation question about the same quality change scenario. The quality change is described in terms of fish variety and catch, bacteria levels and algae blooms, water odor and color, and clarity of the lake.

A. Data Set Restrictions

Of the 626 returned surveys, 44 respondents did not answer the trip questions or the CV question and were therefore discarded. The visitors who reported unusually large travel distance or excessive reported trips were also excluded. This was done by limiting the travel time one way to 5 hours (20 surveys discarded) and limiting the number of total trips to 52 (34 discarded), allowing one trip per weekend.

As HHW did, visitors were also excluded for reporting fewer trips under the improved water quality than they stated for the previous year. A surprisingly large number of surveys, 145, were discarded due to this restriction; a loss of 27.5% of the remaining surveys. This large loss of observations deserves further discussion.

---

[5] See appendix 1 for a copy of the survey.

Maybe the visitors unintentionally reported Plan B trips less than past trips. The respondents were asked for their total trips over the previous year as the first question in the survey. Before being asked the contingent trips under Plan B they were given a description of the current water quality conditions. I think even the current conditions came as a surprise to many of the respondents by making them aware of risks of "algae blooms" and other descriptions they may have never considered before. I think maybe the respondent is reacting to the "current conditions" description. Had they known this information before their trips over the last year, they actually would have gone less.

After the initial water conditions were described, the survey contained three different quality plans. Plan A depicted the Lake if nothing was done, showing significantly deteriorated water quality. As reported by Azevedo, Herriges, and Kling (2001) the contingent trips plunged under this plan. The next plan was Plan B, the proposed water quality improvement plan.

Therefore a possible additional explanation is, Plan A preceding Plan B biases the answers to Plan B. Maybe the reason the respondents are not being careful with their answers is they report Plan B trips as a significant increase over Plan A, but they are not considering what their original reported trips were and thus Plan B trips is significantly greater than Plan A trips but still actually less than past trips. As evidence I checked the number of visitors who put *more* trips under Plan A, the *degraded* water quality. Only 4.2% of the visitors put *more* trips compared to 27.5% who failed the quality consistency condition under Plan B. Maybe this is evidence Plan A being first biases the trips reported for Plan B.

I mention this issue simply as a curiosity and something to ponder when designing future surveys. A few suggestions I would recommend would be: 1) The first question of the survey asked the visitors to report their trips over the last year but they reported their trips for each of the four seasons and never were asked to total the number of trips. Maybe it would be better to ask the respondent to also total this number so they have total trips in mind. 2) When the respondent reports their trips, ask them to go to a specified page further along in the survey and record that number again in a place right before they are asked to report their contingent trips for each alternative plan. Although, I know it is not ideal to have respondents flipping ahead in the survey, probably a better solution is to only ask for, "additional trips" under the improved water quality. Then the visitor has no choice but to leave trips unchanged or report higher trips. 3) Ask how many trips they would have taken over the past year if they had been fully aware of the "current conditions" of the lake. 4) Or, the survey also asked their expected trips next year, but it too preceded the description of current conditions of the lake. Another alternative is to have this question asked after the description.

B. Summary Statistics

Summary statistics of the data set are given in Table 1. The average number of observed behavior trips is 10.9 and the average number of contingent behavior trips is 15.7 (13.9 for low improvement and 17.2 for high improvement). However these averages are inflated due to the on-site sampling. Those who take a high number of trips are more likely to be intercepted and therefore overrepresented in the sample. Modeling techniques will be employed to control for the on-site sampling.

The respondents answered yes to the contingent valuation question 54.0% of the time. The bid values ranged from $45 to $660 with a mean bid value of $333.22. Income was elicited in categories with income levels coded at the midpoints of the income ranges (the upper range was coded as $200,000), the mean income was $62,182. Again, due to the on-site sampling, the data set is skewed to those individuals with a larger income than the population as a whole. The higher income individuals take more trips and are therefore more likely to be intercepted and then overrepresented in the sample.

The statistics presented in Table 1 for the observed behavior trips and expected trips indicate the visitors, on average, reported expected trips next year to be basically the same as the trips they reported over the last year. A simple test of this hypothesis is done by randomly pairing the reported values of observed behavior trips and expected trips, then taking the difference between the two. This difference is treated as a random variable distributed normally. The null hypothesis mean observed behavior trips equals mean expected trips cannot be rejected at the 0.10 significance level.[6]

Table 1. Summary Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
Travel Cost (p)	$63.30	$58.13	0	$512.50
Observed Behavior Trips (x)	10.91	11.65	1	50
Expected Trips	11.26	12.15	0	50
Contingent Behavior Trips (x^*)	15.74	16.03	1	100
Income (m)	$62,182	$38,373	$7,500	$200,000
B Bid (B)	$333.22	$136.75	$45	$660
Yes	.54	.5	0	1
D	.55	.49	0	1

Sample Size=383

[6] To test the null hypothesis mean past trips equals mean expected trips, I also performed the nonparametric signed-rank test as Huang, Haab, and Whitehead (1997) did. In contrast to the parametric test I present in this paper, the null hypothesis is rejected at the 0.01 significance level. However, I feel the signed-rank test is not as valid for my sample due to the large number of zero differences between past trips and expected trips (34.1% of the respondents reported expected trips to be the same as past trips). The signed-rank test discards the zero differences and only tests the remaining pairs of values.

HHW concluded expected trips should be combined with contingent valuation data. However they had a significant difference between observed behavior trips and expected trips. Also, the Clear Lake survey specifically asks the visitor to consider their trips "over the last year" making it clear the comparison is with observed behavior trips. For these reasons this analysis is only with observed behavior trips as the independent variable.

V. Estimation Results

The travel cost model corrected for on-site sampling is first estimated and then the trip change function and the WTP variation function are estimated separately, and lastly the joint model corrected for on-site sampling.

A. Independently Estimated Models

The maximum likelihood coefficients from the travel cost model have the appropriate qualitative signs with recreation trips inversely related to price and increasing with income (table 2). All coefficients are significant at 1% level except the constant for observed behavior trips. The estimated average demand curve for the contingent behavior trips shifts out with the improvement in water quality causing an increase in the consumer surplus estimate. The contingent behavior trips are also less responsive to price and income.

Annual consumer surplus estimates from the count data recreation demand models are easily estimated if the count regression model uses the mean exponential function (i.e. $\lambda_1 = \exp\left(\theta' z_i\right)$. Consumer surplus is the area under the aggregate demand curve from the beginning price $\left(p^B\right)$ to the choke price $\left(p^C\right)$. Since at the choke price demand is zero and at the beginning price demand is the observed number of trips, annual consumer surplus for each individual is

$$CS_i = \int_{P^B}^{P^C} \lambda_i(P)\,dP$$
$$= \frac{-\lambda_i}{\beta^P}$$

where β^P is the coefficient for the price variable and λ_i is the predicted number of trips.

The average predicted trips for the observed behavior data is 3.18 leading to annual average consumer surplus estimates of $205.78 per individual given current conditions. The average predicted contingent behavior trips is 6.05, averaging the predicted trips for the moderate improvement, 5.43, and the high improvement, 6.54, leading to an annual mean consumer surplus of $486.73.

The above estimates are calculated using coefficients corrected for on-site sampling, however, as discussed, the independent variables themselves are also affected by administering an intercept survey. To obtain fitted population trips corrected for population characteristics, denoted as x^p and x^{*p}, requires using population averages for the independent variables. I assume the population is the state of Iowa, as done in the first essay of this dissertation. The bottom of table 2 lists the fitted population trips corrected for population characteristics and the resulting annual consumer surplus estimates.

Table 2. Recreation Demand: BVPLN
(Standard Errors in Parentheses)[a]

Parameter	x	x^*
Constant	0.18 (0.14)	1.07* (0.10)
Travel Cost (p)	-1.55* (0.10)	-1.25* (0.07)
Income	1.68* (0.14)	1.16* (0.11)
D		0.19* (0.04)
σ	1.2* (0.05)	0.97* (0.03)
ρ	0.98* (0.01)	
Consumer Surplus per Trip	64.63* (4.06)	80.42* (4.79)
Fitted Population Trips	3.18 (7.42)	6.05 (8.87)
Corresponding Annual Consumer Surplus	205.78 (12.91)	486.73 (28.98)

	x^p	x^{*p}
Fitted Population Trips (corrected for population characteristics)	0.37 (0.90)	1.26 (1.93)
Corresponding Annual Consumer Surplus	23.79 (1.49)	101.20 (6.03)

* Significant at 1% level.

[a] The travel cost coefficient is scaled by 100, and the income coefficient is scaled by 100,000.

Table 3. Independent and Joint Estimation
(Standard Errors in Parentheses)

Parameter	Trip Change Fn.	WTP Variation Fn.	Joint Model
α		-151.02 (314.96)	317.45 (228.98)
β^{cb}	-20.12* (3.02)		-24.99* (3.38)
β^{cv}		-2.05 (2.10)	-0.87 (1.90)
μ^{cb}	1.11* (0.06)		1.14* (0.07)
μ^{cv}		1.01* (0.01)	1.01* (0.01)
D	-4.76* (1.28)	187.21 (143.16)	
σ_1		1086.06 (754.33)	1298.77 (830.04)
σ_2	7.45* (0.28)		7.61* (0.28)
ρ			-0.19* (0.06)
WTP		473.76 (366.87)	989.50 (424.91)
WTP^{p}		222.19 (523.97)	773.62 (523.47)

* Significant at 1% level.

In Table 3, the maximum likelihood estimates from the trip change function are all of the proper qualitative sign and significant at the 1% level. Only the income coefficient is significant from the WTP variation function, leading to a large standard error for the WTP estimate.

B. Use and Nonuse Values

A second approach for estimating the WTP for the water quality improvement scenarios can be calculated as the difference in the annual consumer surplus estimates from the recreation demand models, $\hat{w}^{CB}$. It is labeled with CB since it uses the contingent behavior data. However, $\hat{w}^{CB}$ only measures use value. The WTP estimate from the contingent valuation question, $\hat{w}^{CV}$, includes both use and nonuse values. If one assumes the weak complementarity condition holds for all the visitors, then this distinction is moot, and the two stated preference data can be assumed to measure the same underlying preferences. But with many environmental amenities, measuring the nonuse value can be a significant portion of the total welfare. It is possible to separate the WTP estimates into use and nonuse values by subtracting $\hat{w}^{CB}$ from $\hat{w}^{CV}$:

$$\hat{w}^{CB} = \$101.20 - \$23.79 = \$77.40$$
$$\hat{w}^{CV} = \$222.19$$

resulting in an estimated nonuse value of Clear Lake at $144.79; 65.2% of the total.

Notice, the nonuse value was estimated using the fitted values with the population averages used as the independent variables. While the WTP estimate from the contingent valuation data has been adjusted with respect to the independent variables, the estimated coefficients are still uncorrected. The WTP estimate from the contingent behavior data has corrected the coefficients and the independent variables for on-site sampling, and to do the same with the contingent valuation data requires a joint model to make it possible to reweight the density according to that which is truncated and endogenously stratified, observed trips.

The next section discusses the estimates from the joint model that does control for the on-site sampling using both the contingent behavior and contingent valuation data.

C. Jointly Estimated Model

The estimates from the joint model mimic the independent models with the coefficients from the trip change function all being significant at the 0.01 level and only the income coefficient being significant at that level from the WTP variation function (Table 3). Unfortunately, the joint model with the dummy variable for the medium and high improvement water quality plans would not converge. Therefore, this dummy variable was excluded, and the reported WTP estimate is for the average of the two plans. Also, the null hypothesis, $\beta^{cb} = \beta^{cv}$ and $\mu^{cb} = \mu^{cv}$, is rejected based on a likelihood ratio test, and therefore this specification is excluded from the analysis.

Surprisingly $\hat{\rho}$, the correlation coefficient, is significantly estimated as a negative number, meaning as the visitors take less trips their estimated WTP from the variation function increases. Since the correction for on-site sampling essentially leads to more weight given to the low trip takers, the joint WTP estimate is significantly larger than the WTP estimate from the variation function separately. However, the standard errors on the WTP estimates are large indicating the model does not have much explanatory power. A positive correlation between additional trips taken and WTP was expected leading to a lower joint WTP estimate.

It appears the Clear Lake sample is not well suited to this modeling strategy. Adding the price of the recreation trip as an explanatory variable in the WTP variation function and in the joint model produces insignificant coefficients for the price. The respondents to the

Clear Lake survey who take many trips (i.e. have a lower price on average) are not significantly more likely to answer "yes" to the contingent valuation question.

VI. Conclusions

The Clear Lake data set is a rich data set asking the visitors revealed preference and two stated preference questions, contingent behavior and contingent valuation. This paper has discussed ways of utilizing this data to measure welfare gains for a water quality improvement plan. In particular, estimates of the welfare gains were derived in three ways; contingent behavior and contingent valuation separately, and an approach to jointly model the data. The joint approach in this paper is unique since it combines three data sources (one RP and two SP) instead of the usual two (one RP and one SP). This is done by exploiting consumer welfare theory to derive a trip change function that includes both past trips (RP) and plan B trips (SP) in one function along with the WTP variation function.

In addition, this paper has shown how to correct WTP estimates from contingent valuation data for on-site sampling. The approach is to jointly model the contingent valuation data with the trip data and then reweight the joint distribution appropriately. Surprisingly, the correlation between recreation trips and WTP for a quality improvement was found to be negative leading to unexpected results of increased WTP estimates when correcting for on-site sampling. However the standard errors on the WTP estimates is large indicating the model does not have much explanatory power, and other applications of this model may lead to more significant results.

VII. References

Azevedo, C.D. "Linking Revealed and Stated Preference Data in Recreation Demand Modeling." 1999. Ph.D. dissertation. Ames, IA: Iowa State University.

Azevedo, C.D., J.A. Herriges, and C.L. Kling. 2000. "Ask a Hypothetical Question, Get a Valuable Answer?" Working Paper

Azevedo, C.D., J.A. Herriges, and C.L. Kling. 2001. "Valuing Preservation and Improvements of Water Quality in Clear Lake." Staff Report 01-SR 94.

Cameron, T. 1988. "A New Paradigm for Valuing Non-market Goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression." *Journal of Environmental Economics and Management* 15:355-79.

Cameron, T. 1992. "Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods." *Land Economics* 68 (August): 302-17.

Cesario, F.J. 1976. "Value of Time in Recreation Benefit Studies." *Land Economics* 52: 32-41.

Huang, J., T. Haab, and J. Whitehead. 1997. "Willingness to Pay for Quality Improvements: Should Revealed and Stated Preference Data be Combined?" *Journal of Environmental Economics and Management* 34(3) (November): 240-55.

Layman, R.C., J.R. Boyce, and K.R. Criddle. 1996. "Economic Valuation of the Chinook Salmon Sport Fishery of the Gulkana River, Alaska, under Current and Alternate Management Plans." *Land Economics* 72(1) (February): 113-28.

Shaw, D. (1988). "On-site Samples' Regression: Problems of Non-Negative Integers, Truncation, and Endogenous Stratification." *Journal of Econometrics* 37: 211-23.

Whitehead, J.C. 1995. "Willingness to Pay for Quality Improvements; Comparative Statics and Interpretation of Contingent Valuation Results." *Land Economics* 71(2) (May): 207-15.

CHAPTER 4. RECREATION DEMAND USING PHYSICAL WATER QUALITY MEASURES

A paper to be submitted to a journal in the field

Kevin J. Egan[1,2], Joseph A. Herriges[1], Catherine L. Kling[1], John A. Downing[3]

I. Introduction

According to the U.S. Environmental Protection Agency (EPA) (2003), significant strides have been made in reducing the impacts of point source pollutants on our aquatic resources. However, our waters continue to remain impaired, "primarily due to complex pollution problems caused by nonpoint source pollution (p. 1-1)." This report continues stating that the most recent (2000) national water quality inventory shows 45% of assessed lake acres are impaired. Two leading causes of these impairments are nutrients and siltation; with nutrients alone and the related biological growth creating approximately half of the assessed impaired waters (EPA, 2000). In states like Iowa, agriculture is a primary source of nutrients, though urban runoff also contributes. Iowa's impaired waters list reports nutrients and suspended solids as practically the sole source of the impairment (EPA Water Quality Inventory for the State of Iowa, 2003).[4]

Therefore, an important empirical question is if, or to what degree, do visitors consider the physical water quality (i.e. the data limnologists collect when studying lakes) of an aquatic resource when making recreation choices? Specifically, are they responsive to physical water quality measures such as nutrients, or are other lake characteristics more important, for example, location, or available facilities? The relationship between physical

---

[1] Predoctoral research associate and Professors, respectively, Department of Economics, Iowa State University.
[2] Primary author.
[3] Professor, Department of Ecology, Evolution, and Organismal Biology, Iowa State University.
[4] Available on the internet, the URL is: http://oaspub.epa.gov/waters/w305b_report.state?p_state=IA

water quality measures and recreational use is also central to understand as these scientific measures are the most objective and quantifiable. In addition, the EPA considers physical water quality measures when determining which lakes are impaired. A lake that is considered impaired becomes a candidate for the total maximum daily load (TMDL) process. A TMDL is a calculation of the greatest amount of a pollutant that a water body can receive without violating water quality standards. The pollutant may be one of the physical water quality measures collected, such as total phosphorus or nitrogen. Therefore, directly including physical water quality measures in the analysis allows welfare calculations to be based on improvements in levels, as stipulated by the TMDL's, to remove the water body from the impaired waters list.

While there is of course an important question regarding the degree to which visitors respond to scientific water quality measures, visitors may not directly respond to the level of nutrients in the water. However, reports by limnologists state, "Increased nutrient supply to fresh waters has been associated with algal blooms, imbalances in water ecosystems, fish kills, increase in toxin-producing microorganisms, and reduced aesthetic value of lakes and streams" (Mallarino et al. 2002, p. 440). Thus to the extent that visitors respond to these ecosystem services, physical water quality measures may predict recreation choices.

The purpose of this paper is to estimate the responsiveness of recreational lake trips to physical measures of water quality collected from the lakes. A few papers have addressed this issue. Feather and Hellerstein (1997) estimated the recreational benefits from the conservation reserve program. The authors included soil erosion as an explanatory variable for recreational trips, theorizing that the conservation reserve program reduces erosion, which they show is correlated with physical water quality measures such as nitrogen and phosphorous. The final link needed is the physical water quality measures effect on recreational behavior. Feather and Hellerstein acknowledge that this information is unknown, but they assume the relationship exists to complete the argument for erosions effect on recreation behavior.

More recently, two papers have estimated the responsiveness of recreation behavior to a few measures of physical water quality. Phaneuf, Herriges, and Kling (2000) estimate a Kuhn-Tucker model analyzing angler behavior in the Great Lakes. They include catch rates for particular fish species of interest as well as a toxin measure derived from the average toxin levels (ng/kg-fish) given in a study by De Vault et al. (1989). The authors state that the toxin level, a measure of the presence of environmental contaminants, is likely to influence the recreation decision much in the same way, in this paper, we expect physical measures of water quality like nutrients will affect recreation decisions. The second paper is Von Haefen (2003) who uses two of the same physical water quality measures as this paper, total phosphorus and secchi depth.

All of these papers find significant effects for their included quality variables, even with limited numbers of observations (Von Haefen) or aggregated sites (Phaneuf et al., Feather and Hellerstein, and Von Haefen). This paper extends this line of research analyzing a comprehensive data set in terms of its expansive recreation behavior and physical water quality collected. The Iowa State University Limnology laboratory, led by Dr. John Downing, a limnologist at Iowa State University, is conducting a 5 year study of 129 of Iowa's principal lakes. To complement this data, a random population survey was sent to 8,000 Iowans to collect information on their recreation behavior to all of the 129 lakes.

Iowa is a unique setting for this analysis as the lake destinations are one of the primary recreational activities available in Iowa. Survey results suggest that Iowans averaged 9.7 total trips for 2002 to Iowa lakes.[5] In addition the water quality in Iowa's lakes varies from a few clean lakes with up to 15 feet of visibility to other lakes having some of the highest concentrations of nutrients in the world. On average the water quality is poor, as evidenced by 31 of the 129 principal lakes officially listed as impaired by the EPA.

We employ the repeated Mixed Logit random utility framework first introduced by Revelt and Train (1998), and in the area of recreation demand by Train (1998). More recently in recreation demand, Herriges and Phaneuf (2002) utilize the error components interpretation of Mixed Logit, while Von Haefen (2003) follows Train (1998) employing the random parameters interpretation. In this paper we utilize the random parameters interpretation to model recreation behavior to Iowa's lakes. The Mixed Logit model is a flexible structure allowing the analyst to most appropriately model recreationist's behavior by incorporating the substitution and correlation patterns between various lakes.

This paper illustrates that visitor's trip behavior is significantly responsive to physical water quality measures. WTP estimates are calculated from three different scenarios. The first scenario improves the water quality of all 129 lakes to equal the water quality of West Okoboji Lake, one of the cleanest lakes in Iowa. The second scenario is a less ambitious, more realistic plan which improves nine lakes evenly placed throughout the state to equal the

---

[5] This number includes single day and multiple day trips to the 129 principal lakes included in this analysis as well as total trips reported in the "other Iowa lakes" category. This number also averages the results from the mail survey and a follow-up telephone survey administered to the mail survey non-respondents. The concern was the mail survey non-respondents may be on average less avid recreators, as is the case, with this group averaging slightly more than half as many trips as the mail survey respondents. However, only total trips were collected in the telephone survey and in this paper we only use single day trips. Therefore, in the rest of the paper only the mail survey respondents' single day trips to the 129 principal lakes are analyzed.

water quality level of West Okoboji Lake. The last scenario considers improving the 31 impaired lakes as listed by the EPA to a high enough quality level to remove them from the list. These scenarios show that Iowans highly value their lakes, but would benefit the most from a few more lakes with superior water quality rather than all recreational lakes being brought to an adequate water quality level.

II. Mixed Logit Model

The Mixed Logit model was chosen since it exhibits many desirable properties including, "it allows for corner solutions, integrates the site selection and participation decisions in a utility consistent framework, and controls for the count nature of recreation demand (Herriges and Phaneuf, 2002)."

Assume the utility of individual i choosing site j on choice occasion t is of the form

$$U_{ijt} = V\left(X_{ij};\beta_i\right) + \varepsilon_{ijt},\ i=1,\dots,N;\ j=0,\dots,J;\ t=1,\dots,T \tag{1}$$

where V represents the observable portion of utility, and from the perspective of the researcher, ε_{ijt}, represents the unobservable portion of utility. A mixed logit model is defined as the integration of the logit formula over the distribution of unobserved random parameters (Revelt and Train, 1998). If the random parameters, β_i, were known then the probability of observing individual i choosing alternative j on choice occasion t would follow the standard logit form

$$L_{ijt}\left(\beta_i\right) = \frac{\exp\left(V_{ijt}\left(\beta_i\right)\right)}{\sum_{k=0}^{J}\exp\left[V_{ijt}\left(\beta_i\right)\right]}. \tag{2}$$

Since the β_i's are unknown, the corresponding unconditional probability, $P_{ijt}(\theta)$, is obtained by integrating over an assumed probability density function for the β_i's. The unconditional probability is now a function of θ, where θ represents the estimated moments of the random parameters. This repeated Mixed Logit model assumes the random parameters are *i.i.d.* distributed over the individuals so that

$$P_{ijt} = \int L_{ijt}(\beta) f(\beta \mid \theta) d\beta. \tag{3}$$

No closed form solution exists for this unconditional probability and therefore simulation is required for the maximum likelihood estimates of θ.[6, 7]

Following Herriges and Phaneuf (2002), a dummy variable, D_j, is included which equals one for all of the one through J recreation alternatives and equals zero for the stay-at-home option *(j=0)*. Including the stay-at-home option allows a complete set of choices, including in the population those individuals who always "stay at home" on every choice occasion and do not visit any of the sites. It is convenient to partition the individual's utility into the stay-at-home option or choosing one of the J sites

$$U_{ijt} = \begin{cases} \beta^{z'} z_i + \varepsilon_{i0t} \\ \beta_i' x_{ij} + \alpha_i + \varepsilon_{ijt}, \quad j = 1, \ldots, J \end{cases} \tag{4}$$

where α_i is the random parameter on the dummy variable, D_j, which does not appear since it equals one for $j = 1, \ldots, J$ and zero for $j = 0$. The vector z_i contains socio-demographic data such as income and age, and x_{ij} represents the site characteristics that vary across the

---

[6] Train (2003) describes simulation methods for use with mixed logit models, in particular maximum simulated likelihood which we employ. Software written in GAUSS to estimate mixed logit models is available from Train's home page at http://elsa.berkeley.edu/~train.

[7] As in the first essay of this dissertation, randomly shifted and shuffled uniform draws are used in the simulation process (Hess, Train, and Polak, 2003). The number of draws used in the simulation was 750.

lakes including attributes such as facilities at the lake as well as water quality measures. Notice the parameters associated with the socio-demographic data are not random as this information does not vary across the sites.[8]

The random coefficient vectors for each individual, β_i and α_i, can be expressed as the sum of population means, b and a, and individual deviation from the means, δ_i and γ_i, which represents the individual's tastes relative to the average tastes in the population (Train, 1998). Therefore redefine

$$\beta_i' x_{ij} = b' x_{ij} + \delta_i' x_{ij}, \tag{5}$$

$$\alpha_i = a + \gamma_i, \tag{6}$$

and then the partitioned utility is

$$U_{ijt} = \begin{cases} \beta^{z'} z_i + \eta_{i0t} & \\ b' x_{ij} + a + \eta_{ijt}, & j = 1,\dots,J \end{cases} \tag{7}$$

where

$$\eta_{ijt} = \begin{cases} \varepsilon_{i0t} & i = 1,\dots,N;\ t = 1,\dots,T \\ \delta_i' x_{ij} + \gamma_i + \varepsilon_{ijt} & j = 1,\dots,J;\ i = 1,\dots,N;\ t = 1,\dots,T \end{cases} \tag{8}$$

is the unobserved portion of utility. This unobserved portion is correlated over sites and trips due to the common influence of the terms, δ_i' and γ_i which vary over individuals. For example, an individual who chooses the stay-at-home option for all choice occasions would have a negative deviation from a, the mean of α_i, while someone who takes many trips would have a positive deviation from a, allowing the marginal effect to vary across individuals. However the parameters do not vary over sites or choice occasions; thus, the

[8] It is possible to interact the socio-demographic data with the sites, if one believed for example that income would affect which lake was chosen.

same preferences are used by the individual to evaluate each site at each time period. Since the unobserved potion of utility is correlated over sites and trips, the familiar IIA assumption does not apply for mixed logit models.

III. Data

The random population sample was obtained from a mail survey sent to 8,000 Iowans in November of 2002.[9] The survey collected trip data for 2001 and 2002 actual trips to 129 lakes as well as anticipated trips for the 2003 season. Of the 8,000 mailed surveys, 882 were undeliverable. A total of 4,423 surveys were returned resulting in a 62% response rate.

The final sample of 3,859 individuals was obtained as follows. Those individuals who returned the survey from out of state were excluded (38 observations). It is impossible to know if these respondents have permanently left the state or reside elsewhere for part of the year. They are excluded since their travel cost calculations could be unrealistically high. Also, those individuals who did not complete the trip questions or did not give a number (i.e. they put a check mark) were excluded (224 observations). Lastly, anyone reporting more than 52 total single day trips to the 129 lakes were excluded (133 observations). Only single day trips are included to avoid the complexity of modeling multiple day visits. Defining the number of choice occasions as 52, allows one trip to one of the 129 Iowa lakes per week. The choice of 52 is arbitrary, but it seems a reasonable cut-off for the total number of allowable single day trips for the season. Invariably some of the respondents who recorded trips greater than 52 did in fact take this number, but since this survey was randomly sent out to Iowans, some of the recipients live on a lake, and it may be those individuals who record

[9] See appendix 2 for a copy of the survey.

hundreds of "trips" simply by returning to their residency. The choice of 52 eliminated about 3% of the returned surveys.[10]

Due to the large number of respondents, we randomly divide the sample into three segments; specification, estimation, and prediction portions. The analysis reported here comes from the specification stage using 1,286 observations. Once the estimation stage is reached the results will be free from any form of pretest bias and the standard errors will be unbiased by the extensive specification search.

IV. Model Application

Respondent's attitudes regarding lake quality as well as socio-demographic data were solicited in the survey instrument.[11] One question asked the respondents to rank, using a total of 100 points, which factors were most important in choosing a lake for recreation. The top three choices were water quality (33 points), proximity (22 points) and park facilities (18 points); all characteristics included in our model. The next largest category was "location of friends/relatives" at 11 importance points. This category, along with the other 17 importance points not mentioned, are not possible to be included in this analysis and will be relegated to the error term. However 72% of the importance points are captured with water quality being the most important, indicating that the respondents do consider the water quality of the lake when making their recreation decisions.

We model the utility individual i receives from choosing lake j on choice occasion t as

[10] A model with 150 choice occasions was also estimated. None of the coefficients from this model change qualitatively from the results presented in this essay except two of the socio-demographic coefficients. The conclusions from the water quality scenarios discussed later in the essay are also unchanged.
[11] See Azevedo et al. (2003) for a summary report of the results from the survey.

$$U_{ijt} = \begin{cases} \beta^{z'} z_i + \varepsilon_{i0t} \\ -\beta^p P_{ij} + \beta^{q'} Q_j + \beta_i^{a'} A_j + \alpha_i + \varepsilon_{ijt}, \quad j = 1, \ldots, J \end{cases} \tag{9}$$

where z_i is the socio-demographic data summarized in table 1, P_{ij} is the travel cost from each Iowan's residency to each of the 129 lakes, as calculated with *PCMiler*.[12] One component of the price is the out-of-pocket cost computed as the roundtrip travel distance multiplied by $0.25 per mile. The other component is the opportunity cost of time calculated as one-third the estimated roundtrip travel time multiplied by the respondents wage rate (calculated as the respondents reported income divided by 2000). The vector Q_j denotes the physical water quality measures collected by John Downing's team and A_j represents the attributes of the lake. As shown in equation (9), notice that the parameters on the lake attributes and the dummy variable, D_j, are random. These six variables are assumed to be independently normally distributed with the mean and dispersion of each variable estimated.

Table 1 lists the summary statistics for trips and the socio-demographic data. The average number of total single day trips for all 129 lakes is 6.68 varying from some respondents taking zero trips and others taking 52 trips. The survey respondents are more likely to be older, male, have a higher income, and more educated than the general population, but this overrepresentation is less severe than in the first essay of this dissertation when the sample was collected on-site. Schooling is entered as a dummy variable equaling one if the individual has attended or completed some level of post high school education.

---

[12] *PCMiler* is a product of ALK Technologies, Inc (2003) and is a software package designed for use in the transportation and logistics industry. Specifically we used the PC*Miler|Streets version 17 software with BatchPro.

The summary statistics for trips, the physical water quality measures, and the lake attributes are listed in table 2. The sample size is the 129 lakes. The average trips per lake is 0.05 with the maximum value equaling 0.50. Since there are about 1.2 million households in Iowa this means that the average lake receives about 60,000 trips annually and the highest visited lake, Saylorville Lake, receives about 600,000 annual trips. The average price of a recreational trip to a lake is $135.79, although more meaningfully the average price of a lake visited is $85.09. The lakes in the corner of the state will have higher average travel costs as most of the state residents would have to travel further to get there. The size of the lakes varies considerably from 10 acres to 19,000 acres. Thus, the log of acres is used in the estimation.

Table 1. Socio-demographic Summary Statistics

Variable	Mean	Std. Dev.	Minimum	Maximum
Total Day Trips	6.68	10.46	0	52
Income	$56,140.52	$37,436.48	$7,500	$200,000
Male	0.67	0.46	0	1
Age	53.36	16.47	15	82
School	0.66	0.47	0	1
Household Size	2.61	1.32	1	12

Sample Size=1,286 individuals

Four dummy variables are included to capture different amenities at each lake. The first is a "ramp" dummy variable which equals one if the lake has a cement ramp as opposed to a gravel ramp or no boat ramp at all. The second is a "wake" dummy variable which equals one if wakes are allowed and zero otherwise. About 66% of the lakes allow wakes and therefore 34% of lakes are "no wake" lakes. The "state park" dummy variable equals one if the lake is located in a state park, true for 38.8% of the lakes. The last dummy variable is the "facilities" dummy variable. This information, as the rest of the dummy variables, was taken from the "Fishing Guide For Iowa Lakes" published by the Iowa Department of Natural Resources. This report divides all Iowa Lakes into those with "accessible facilities" and those without. Accessible facilities include things like restrooms, picnic tables, or vending machines. A concern may be that facilities would be strongly correlated with the state park dummy variable. It turns out there is enough variation between the two to warrant including both. 50 lakes are located in state parks and 50 lakes have accessible facilities, but only 26 of these 50 lakes have both.

Table 2. Lake Characteristics & Water Quality Summary Statistics

Variable	Mean	Std. Dev.	Minimum	Maximum
Avg. Trips per Lake	0.052	0.08	0	0.504
Price	135.79	29.47	94.12	239.30
Acres	672.20	2,120.30	10	19,000
Log (Acres)	4.81	1.69	2.30	9.85
Ramp	0.86	0.35	0	1
Wake	0.66	0.47	0	1
State Park	0.39	0.49	0	1
Facilities	0.39	0.49	0	1
Secchi Depth (m)	1.17	0.92	0.09	5.67
Chlorophyll (ug/l)	40.93	38.02	2.45	182.92
NH_3+NH_4 (ug/l)	292.15	158.57	72	955.34
NO_3+NO_2 (mg/l)	1.20	2.54	0.07	14.13
Total Nitrogen (mg/l)	2.20	2.52	0.55	13.37
Total Phosphorus (ug/l)	105.65	80.61	17.10	452.55
Silicon (mg/l)	4.56	3.24	0.95	16.31
pH	8.50	0.33	7.76	10.03
Alkalinity (mg/l)	141.80	40.98	73.83	286.17
Inorganic SS (mg/l)	9.43	17.87	0.57	177.60
Volatile SS (mg/l)	9.35	7.93	1.64	49.87

Sample Size=129 lakes

This analysis includes several physical water quality measures collected by John Downing and his team. Table 2 lists the included physical water quality measures. Chlorophyll is an indicator of plant biomass or algae and leads to greenness in the water. Three nitrogen levels are included, with the NH_3+NH_4 measuring particular types of nitrogen such as ammonia which can be toxic. NO_3+NO_2 measures the nitrates in the water, and lastly total nitrogen is included in units of milligrams per liter. Total phosphorous is usually the principal limiting nutrient in Iowa lakes, meaning it most likely determines algae growth. Silicon is important to diatoms which extract it from the water to use as a component of their cell walls. Diatoms, in turn, are a key food source for marine organisms. The acidity of the water is measured by "pH" with levels below 6 or above 8 indicating unhealthy lakes. As table 2 notes, all of the pH levels in this sample are tightly dispersed between 7.3 and 10. This term is included as a quadratic variable to reflect that low or high values are detrimental to water quality, but since no low values are observed, a different functional form for pH may be more appropriate. Alkalinity is the concentration of calcium or calcium carbonate in the water. Plants need carbon to grow and all carbon comes from alkalinity, therefore alkalinity is an indication of the abundance of plant life. ISS is the inorganic suspended solids, basically soil and silt in the water due to erosion. VSS, is volatile or organic suspended solids, both measures that will decrease clarity in the water.

EPA's, "Nutrient Criteria Technical Guidance Manual (2000)," states the four paramount variables for nutrient criteria are total phosphorus, total nitrogen, chlorophyll, and Secchi depth. Downing considers inorganic suspended solids and organic suspended solids to be crucial indicators as well. For these reasons, model A, contains this set of six physical water quality measures. A second model, model B, includes the complete list of eleven water quality measures. Estimating two models allows us to observe the stability of the parameters across different specifications.

Now turning to the levels of the physical water quality measures, it is evident that considerable variation is present across the lakes. For example, secchi depth varies from a low of 0.09 meters to a high of 5.67 meters and total phosphorus varies from 17 ug/L to 453, some of the highest concentrations in the world according to Downing. All of the physical water quality measures are the average values for the 2002 season. Samples were taken from each lake three times throughout the year, in Spring/early Summer, mid-Summer, and late Summer/Fall to include seasonal variation.[13]

V. Results

The results for Model A and B are divided into two tables, 3a and 3b. For both models, the coefficients for the socio-demographic data, price, and the random coefficients on the amenities and α are given in table 3a. Table 3b lists the coefficients for the physical water quality measures for both models. All of the coefficients are significant at the 1% level except for a few of the socio-demographic data. For model B, with eleven physical water quality measures, only the "male" dummy variable is not significant. In Model A, income, household size, and the quadratic term on age are insignificant. Note that the socio-demographic data was included in the conditional indirect utility for the stay-at-home option. Therefore, the negative income coefficient indicates that as income rises the respondents are less likely to stay at home and more likely to visit a lake (i.e. lake visits are a normal good). Males, higher educated, and larger households are all more likely to take a trip to a lake.

---

[13] The Iowa State University's Limnology Laboratory has a website for the Iowa Lakes Survey Project. The URL is: http://limnology.eeob.iastate.edu/IowaLakesSurvey.aspx where you will find an outline of the project and complete results to date.

Unlike the models in the first essay in this dissertation, age has a convex relationship with the stay-at-home option and therefore a concave relationship with trips. For Model B, the peak occurs at about age 37, which is consistent with the estimate of larger households taking more trips, as at this age the household is more likely to include children.

Table 3a. Repeated Mixed Logit Estimates (Standard Errors in Parentheses)[a]

Parameter	Model A: 6 Physical WQ Measures		Model B: 11 Physical WQ Measures	
	Mean	Dispersion	Mean	Dispersion
Income	-0.008		-0.12*	
	(.007)		(0.007)	
Male	-4.98*		-0.31	
	(0.42)		(0.42)	
Age	-0.24*		-0.58*	
	(0.07)		(0.08)	
Age^2	0.0001		0.0078*	
	(0.00006)		(0.0007)	
School	-4.45*		-3.44*	
	(0.40)		(0.40)	
Household	-0.41		-1.24*	
	(0.17)		(0.17)	
Price	-0.17*		-0.17*	
	(0.0006)		(0.0007)	
Log(Acres)	4.60*	3.81*	5.13*	4.05*
	(0.064)	(0.057)	(0.067)	(0.06)
Ramp	11.60*	17.85*	14.87*	18.79*
	(0.78)	(0.51)	(0.89)	(0.59)
Facilities	1.18*	18.09*	3.54*	16.78*
	(0.26)	(0.28)	(0.24)	(0.25)
State Park	8.00*	15.15*	6.67*	13.99*
	(0.26)	(0.27)	(0.24)	(0.27)
Wake	2.76*	15.81*	-1.64*	15.57*
	(0.30)	(0.33)	(0.30)	(0.29)
α	-8.97*	3.01*	-9.19*	3.12*
	(0.05)	(0.04)	(0.05)	(0.04)

*Significant at 1% level.
[a]All of the parameters are scaled by 10, except α (which is unscaled) and the income coefficient (which is scaled by 10,000).

The price coefficient is negative as expected and identical in both models. Now turning to the amenities parameters, again all of the parameters are of the expected sign. As the size of a lake increases, has a cement boat ramp, gains accessible facilities, or is in a state park, on average leads to increased trips. Notice however the large dispersion estimates. For example, in model A the dispersion on the size of the lake indicates 11.1% of the population prefers a smaller lake, possibly someone who enjoys a more private experience. The large dispersion on the "wake" dummy variable seems particularly appropriate given the potentially conflicting interests of anglers and recreational boaters. Anglers would possibly prefer "no wake" lakes and recreational boaters would obviously prefer lakes that allow wakes. It seems the population is almost evenly split with 56.9% preferring a lake that allows wakes and 43.1% preferring a "no wake" lake. Lastly, the mean of α_i is negative indicating that on average the respondents receive higher utility from staying at home, which is not surprising considering the average number of trips is 6.7 out of a possible 52 choice occasions.

The physical water quality coefficients are relatively stable across the two models (table 3b). The only parameter to change qualitatively is total nitrogen. In the model with six included water quality measures, total nitrogen is positive. Downing explains that this is to be expected, given the negative sign on total phosphorus.[14] With such large amounts of phosphorus in the water, more nitrogen can actually be beneficial by allowing a more normal phosphorus to nitrogen ratio. If the ratio becomes too imbalanced more problematic blue-green algae blooms become dominant. Total nitrogen is negative in model B, but two other

[14] All explanations given for the coefficients on the physical water quality measures are my summaries of personal communication with Prof. Downing.

forms of nitrogen are included with the nitrates form (NO_3+NO_2) being positive, possibly for the same reason as just discussed.

Table 3b. Repeated Mixed Logit Estimates (Standard Errors in Parentheses)[a]

Parameter	Model A: 6 Physical WQ Measures	Model B: 11 Physical WQ Measures
Secchi Depth	0.78* (0.05)	0.84* (0.07)
Chlorophyll	0.054* (0.03)	0.06* (0.003)
NH_3+NH_4		-0.002* (0.0006)
NO_3+NO_2		3.16* (0.19)
Total Nitrogen	0.31* (0.01)	-3.21* (0.19)
Total Phosphorus	-0.0033* (0.001)	-0.016* (0.001)
Silicon		0.81* (0.02)
pH		-136.72* (5.83)
pH^2		8.35* (0.34)
Alkalinity		0.038* (0.002)
Inorganic SS	-0.010* (0.008)	-0.089* (0.009)
Volatile SS	-0.18* (0.01)	-0.28* (0.02)
LogLik	-47,740.38	-47,494.17

*Significant at 1% level.
[a]All of the parameters are scaled by 10.

For both models A and B, secchi depth is positive and the suspended solids, both organic and inorganic (volatile), are negative, indicating the respondents strongly value water clarity. However the coefficient on chlorophyll is positive suggesting respondents do not mind some variation of green water. Higher alkalinity acts as a buffering capacity on how much acidity the water can withstand before deteriorating. Therefore, a positive coefficient is consistent with expectations as all of the lakes in the sample are acidic (i.e. pH greater than 7). Silicon is important for diatoms, which in turn are an important food source for marine organisms and therefore a positive coefficient on silicon was expected.

Model B, using eleven physical water quality measures, has pH entered quadratically, as suggested by Downing, reflecting that low or high pH levels are signs of poor water quality. However, in our sample of lakes, all of the pH values are normal or high. The coefficients for pH show a convex relationship (the minimum is reached at a pH of 8.2) to trips, indicating that as the pH level rises above 8.2, trips are predicted to increase. This is opposite of what we expected and further specifications, in consultation with Prof. Downing, will consider this fact.

VI. Water Quality Scenarios

Given the random parameters, β_i, the conditional compensating variation associated with a change in water quality from Q' to Q" for individual *i* on choice occasion *t* is

$$CV_{it}(\beta_i) = \frac{-1}{\beta^p}\left\{ \ln\left[\sum_{j=0}^{J} \exp\left(V_{ijt}(Q'';\beta_i)\right)\right] - \ln\left[\sum_{j=0}^{J} \exp\left(V_{ijt}(Q';\beta_i)\right)\right]\right\}, \tag{10}$$

which is the compensating variation for the standard logit model. The unconditional compensating variation does not have a closed form, but it can be simulated by

$$CV_{it} = \frac{1}{R}\sum_{r=1}^{R}\left\{\frac{-1}{\beta^{p}}\left\{\ln\left[\sum_{j=0}^{J}\exp\left(V_{ijt}\left(Q'';\beta_i^r\right)\right)\right]-\ln\left[\sum_{j=0}^{J}\exp\left(V_{ijt}\left(Q';\beta_i^r\right)\right)\right]\right\}\right\}, \quad (11)$$

where R is the number of draws and r represents a particular draw of β_i from its distribution. The simulation process involves drawing values of β_i and then calculating the resulting compensating variation for each vector of draws, and finally averaging over the results for many draws. Following Von Haefen (2003), 2,500 draws were used in the simulation.

Three water quality improvement scenarios are considered with the results from Model A used for all the scenarios. The first scenario improves all 129 lakes to the physical water quality of West Okoboji Lake, the cleanest lake in the state. Table 4 compares the physical water quality of West Okoboji Lake with the average of the other 128 lakes. All of West Okoboji Lake's measures are considerably improved over the other 128. For example, West Okoboji Lake has slightly over 5 times the water clarity, measured by secchi depth, of the other lakes. Given such a large change, the annual compensating variation estimates of $208.68 for every Iowa household seems reasonable (table 6). Aggregating to the annual value for all Iowans simply involves multiplying by the number of households in Iowa which is 1,153,205.[15] Table 6 also reports the average predicted trips before and after the water quality improvement. Improving all 129 lakes to the physical water quality of West Okoboji Lake leads to a reasonable 14.1% increase in average trips. As expected, the predicted trips to West Okoboji Lake fall by 19.8% from 0.39 average trips per Iowa household to 0.31. Iowans can now choose the nearest lake with the attributes they prefer, instead of traveling further to West Okoboji Lake.

[15] Number of Iowa households as reported by Survey Sampling, Inc., 2003.

Table 4. West Okoboji Lake vs. the other lakes

	West Okoboji Lake	Averages of the other 128 Lakes	Averages of the Nine Zone Lakes
Secchi Dish (m)	5.67	1.13	1.23
Chlorophyll (ug/l)	2.63	41.29	40.13
Total Nitrogen (mg/l)	0.86	2.22	3.64
Total Phosphorous (ug/l)	21.28	106.03	91.11
Inorganic Suspended Solids (mg/l)	1.00	9.49	9.52
Volatile Suspended Solids (mg/l)	1.79	9.43	8.42

The next scenario is a less ambitious, more realistic plan of improving nine lakes to the water quality of West Okoboji Lake (see table 4 for comparison). The state is divided into nine zones with one lake in each zone. Then every Iowan will be within a couple of hours of a lake with superior water quality. The nine lakes were chosen based on recommendations by the Iowa Department of Natural Resources for possible candidates of a clean-up project. The annual compensating variation estimate is $39.71 for each Iowa household. As expected, this estimate is 19.0% of the value if all lakes were improved, even though the scenario involves improving only 7.0% of the lakes. This suggests location of the improved lakes is important and to maximize Iowan's benefit from improving a few lakes, policymakers should consider dispersing them throughout the state.

Table 5. Rathbun Lake vs. the 31 impaired Lakes

	Rathbun Lake	Averages of the 31 Impaired Lakes
Secchi Dish (m)	0.90	0.70
Chlorophyll (ug/l)	6.55	56.76
Total Nitrogen (mg/l)	1.10	2.77
Total Phosphorous (ug/l)	43.87	153.70
Inorganic Suspended Solids (mg/l)	5.42	20.42
Volatile Suspended Solids (mg/l)	3.62	15.49

The last scenario is also a policy oriented improvement. Currently of the 129 lakes, 31 are officially listed on the EPA's impaired waters list. TMDL's are being developed for these lakes and by 2009 the plans must be in place to improve the water quality at these lakes enough to remove them from the list. Therefore, in this scenario the 31 impaired lakes are improved to the physical water quality level of Rathbun Lake, which is just above the threshold of the criteria for listing as impaired. Table 5 compares Rathbun Lake to the averages of the 31 impaired lakes. The table indicates Rathbun Lake seems an appropriate choice with physical water quality measures higher than the averages of the 31 impaired lakes, but much below those of West Okoboji Lake. This scenario is valued considerably lower than the first two water quality improvement scenarios. The estimated compensating variation per Iowa household is $4.87. Consistent with this, the predicted trips only increase 0.3% over the predicted trips with no improvement in water quality.

Table 6. Annual Compensating Variation Estimates using Model A

Average CV	All 129 Lakes Improved to W. Okb.	9 Zonal Lakes Improved to W. Okb.	31 Impaired Lakes Improved to Rathbun
per choice occasion	$4.01	$0.76	$0.09
per Iowa household	$208.68	$39.71	$4.87
for all Iowa households	$240,649,000	$45,788,092	$5,612,219
Predicted Trips (9.80 with current water quality)	11.18	10.06	9.83

A reasonable conclusion is Iowan's have an abundance of lakes at this threshold level, and bringing the low quality lakes up to this level is not much of a benefit. For comparison, the average value from the nine zonal lakes improved to West Okoboji Lake equals $5,087,566 per lake. Therefore, Iowans value about equally one of these nine lakes improved to the superior water quality of West Okoboji Lake over improving 31 lakes to the threshold level of impairment.

VII. Further Research

The large data set allows the methodology of randomly segmenting the sample into specification, estimation, and prediction portions as discussed in Creel and Loomis (1990). The next step will be to complete the specification stage. Some variations include allowing more parameters to be random and entering the log of the physical water quality measures. This research will involve close collaboration with Prof. Downing to insure accurate inclusion of the physical water quality variables, reflecting limnologist's views of this data.

Following completion of the specification stage, the model will be estimated on one-third the data reserved for this purpose. At that point confidence intervals will be constructed for the compensating variation estimates. The confidence intervals as well as the standard errors of the parameters will then be free from any biases due to the specification search. The final step will be out of sample prediction using the final one-third of the data.

Unfortunately, it appears we will not be able to include any information on which of the 129 lakes are good fishing destinations. In personal communication with Jeff Kopaska, from the Iowa Fisheries Bureau, creel surveys are only available for less than 10% of the lakes and even that information is dated. Due to budget cuts no further creel surveys are planned. However, Jeff Kopaska was optimistic in a couple of years the biology division may have fishing data on all 129 lakes that could be included as explanatory variables. This data is untimely for this analysis, but future work may be able to incorporate it.

VIII. Conclusions

The first year survey of the Iowa Lakes Project gathered recreation behavior to 129 of Iowa's principal lakes. This data was combined with extensive physical water quality measures from the same set of lakes gathered by the Iowa State University Limnology Lab. Our analysis employing the repeated mixed logit framework, shows individuals are responsive to physical water quality measures and it is possible to base willingness to pay calculations on improvements in these physical measures. In particular we considered three improvement scenarios, with the results suggesting Iowans more highly value a few lakes with superior water quality rather than all recreational lakes at an adequate level, as determined by being listed as an impaired lake by the Environmental Protection Agency.

A number of important practical findings come directly from this work. Limnologists and other water quality researchers should be interested in the results of this paper, since the general belief is that visitors care about water clarity as measured by secchi depth (how many meters beneath the surface of the water a secchi dish is visible) or water quality in general. However, as stated by Feather and Hellerstein, this link has yet to be demonstrated at least at the individual lake level as done here. By estimating the partial effects of a list of physical measures, we have determined which significantly affect recreationist's behavior. Limnologists and water resource managers can then use this information about what physical lake attributes visitor's trip behavior responds to in designing projects for water quality improvements. Our results indicate water clarity is very important as evidenced by the secchi dish and suspended solids parameters. Also, nutrients in general are found to decrease recreation trips.

The findings from this study also have direct relevance for environmental protection managers and citizens concerned with the water quality in that they can be used to prioritize clean-up activities to generate the greatest recreation benefits for a given expenditure. Not only can the findings be used to determine which lakes and in what order to clean them, but also the most efficient levels of improvement.

IX. References

ALK Technologies, Inc (2003). "PC*Miler|Streets Version 17 User's Guide." ALK Technologies, Inc., Princeton, NJ.

Azevedo, C.D., K.J. Egan, J.A. Herriges, and C.L. Kling (2003). "The Iowa Lakes Valuation Project: Summary and Findings from Year One." CARD report.

Creel, M.D., and J.B. Loomis (1990). "Theoretical and Empirical Advantages of Truncated Count Data Estimators for Analysis of Deer Hunting in California." *American Journal of Agricultural Economics* 72 (May): 434-41.

De Vault, D. S., D. Dunn, P. Bergqvist, K. Wiberg, and C. Rappe (1989). "Polychlorinated Dibenzofurans and Ploychlorinated Dibenzo-p-Dioxins in Great lakes Fish: A Baseline and Interlake Comparison," *Environmental Toxicology and Chemistry* 8:1013-1022.

Feather, P., and D. Hellerstein (1997). "Calibrating Benefit Function Transfer to Assess the Conservation Reserve Program." *American Journal of Agricultural Economics* 79 (February): 151-162.

Herriges, J., and D. Phaneuf (2002). "Introducing Patterns of Correlation and Substitution in Repeated Logit Models of Recreation Demand." *American Journal of Agricultural Economics* 84: 1076-1090.

Hess, S., K. Train, and J. Polak (2003). "On the use of randomly shifted and shuffled uniform vectors in the estimation of a Mixed Logit model for vehicle choice." Working paper.

Iowa Department of Natural Resources (2003). "Fishing Guide For Iowa Lakes." Available at www.iowadnr.com (Retrieved: November 15, 2003).

Mallarino, A.P., B.M. Stewart, J.L. Baker, J.D. Downing, J.E. Sawyer (2002). "Phosphorus indexing for cropland: Overview and Basic Concepts of the Iowa Phosphorus Index." *Journal of Soil and Water Conservation* 57(6): 440-447.

Morey, E., R. Rowe, and M. Watson (1993). "A repeated Nested-Logit Model of Atlantic Salmon Fishing." *American Journal of Agricultural Economics* 75 (August): 578-592.

Phaneuf, D.J., C.L. Kling, and J.A. Herriges (2000). "Estimation and Welfare Calculations in a Generalized Corner Solution Model with an Application to Recreation Demand." *The Review of Economics and Statistics* 82 (1): 83-92.

Revelt, D., and K. Train (1998). "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level." *The Review of Economics and Statistics* 80: 647-57.

Train, K. (1998). "Recreation Demand Models with Taste Differences Over People." *Land Economics* 74 (2) (May): 230-239.

Train, K. (1999). "Mixed Logit Models for Recreation Demand", in: C.L. Kling, J. Herriges (Eds.), Valuing Recreation and the Environment, Edward Elgar Publishing Ltd.

Train, K. (2003). "Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, UK.

U.S. Environmental Protection Agency, Office of Water (4503T) (2003). "National Management Measures for the Control of Nonpoint Pollution from Agriculture." Report EPA-841-B-03-004, U.S. Environmental Protection Agency, Washington, D.C.

U.S. Environmental Protection Agency, Office of Water, Office of Science and Technology (2000). "Nutrient Criteria Technical Guidance Manual: Lakes and Reservoirs." Report EPA-822-B00-001, U.S. Environmental Protection Agency, Washington, D.C.

Von Haefen, R. (2003). "Incorporating observed choice into the construction of welfare measures from random utility models." *Journal of Environmental Economics and Management* 45: 145-65.

CHAPTER 5. GENERAL CONCLUSIONS

I. General Discussion

The first two essays contribute to the recreation demand literature by extending existing models to correct for on-site sampling. The first essay analyzes individual panel data corrected for on-site sampling, and the second essay corrects for on-site sampling when contingent valuation data is jointly modeled with observed and contingent data.

The individual panel data set used in the first essay includes observed behavior trips and contingent behavior trips, contingent on price changes, asked in an intercept survey at Clear Lake in northcentral Iowa. A multivariate mixed Poisson regression model is used to analyze the panel data with a more flexible log-normal distribution used as the mixing distribution, instead of the standard gamma distribution. Using a count data model, the multivariate correction for on-site sampling is a straightforward extension of Shaw's (1988) univariate correction. This essay shows the importance of correcting for on-site sampling, as the adjusted average fitted observed trips and contingent trips decrease by a factor of eleven, resulting in considerably lower annual consumer surplus estimates.

The second essay considers correcting contingent valuation data for on-site sampling. The only way to do this is to jointly model the contingent valuation data with the observed trips which are directly truncated and endogenously stratified due to being collected on-site. This essay extends Huang, Haab, and Whitehead's (1997) analysis by correcting their joint model for on-site sampling. Unfortunately, the Clear Lake data set is not well-suited for this model and future research should consider more flexible functional forms.

The final essay uses two extensive data sets, one economic (4,500 Iowan's trip behavior to 129 of Iowa's principal lakes in 2002), and one ecological (14 physical water quality measures for the 129 Iowa lakes for 2002). The economic data set is from a random population sample sent to 8,000 Iowan's. This essay analyzes Iowan's responsiveness to variation in physical water quality measures. A repeated mixed logit model is employed estimating two models, one with the six most important physical water quality measures included as explanatory variables and one with the full list of physical water quality measures. Both show robust results that Iowan's do consider the physical condition of the lake water when choosing which lakes to visit. In particular, decreased water clarity and increased nutrient concentrations lead to fewer trips.

Lastly, three welfare scenarios were calculated the first improving all 129 principal lakes to a high level of water quality, the second improving nine lakes from around the state to the same high level of water quality, and the last welfare scenario considered improving the impaired lakes (as determined by being listed on the impaired waters listed filed with the EPA) enough to remove them from the impaired waters list. The results indicate Iowan's highly value water quality improvement, but with limited resources, they would prefer a few more lakes with superior water quality over all of the impaired lakes being adequately cleaned.

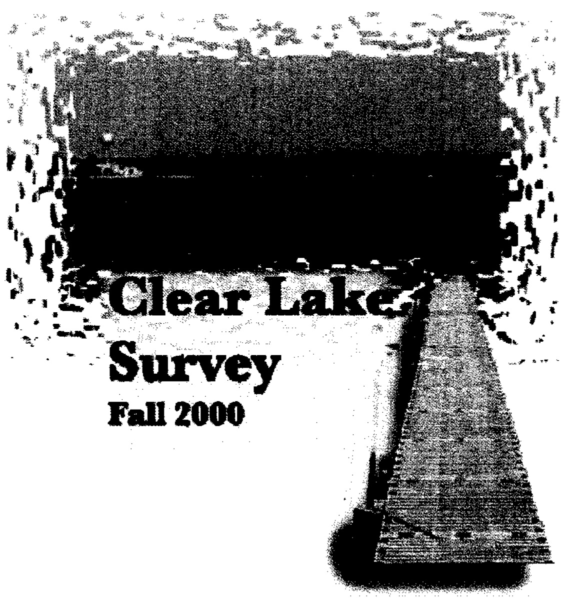
Further research will continue to determine which lakes, in what order, and to what level of clean-up will generate the greatest benefits for a given expenditure. In addition, once total maximum daily load targets are available for the nutrients, the value of achieving the targets can be estimated and ranked; another advantage of estimating welfare values based on physical water quality measures.

II. References

Shaw, D. (1988). "On-site Samples' Regression: Problems of Non-Negative Integers, Truncation, and Endogenous Stratification." *Journal of Econometrics* 37: 211-23.

Huang, J., T. Haab, and J. Whitehead. 1997. "Willingness to Pay for Quality Improvements: Should Revealed and Stated Preference Data be Combined?" *Journal of Environmental Economics and Management* 34(3) (November): 240-55.

APPENDIX 1. CLEAR LAKE SURVEY



Clear Lake Survey

Fall 2000

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

In order to make intelligent decisions concerning the future of Clear Lake, it is important to understand how the lake itself is used, as well as how this use would be affected by possible changes in the quality of the lake. The answers you give to the questions in this survey are very important in this process. Please try to answer each of the questions below. Finally, please keep in mind that, whenever we refer to Clear Lake, we are referring to the lake itself, not the town.

IN THIS FIRST SECTION, WE WOULD LIKE TO ASK YOU ABOUT how you and other members of your household use the body of water known as Clear Lake.

1. For each of the time periods listed below, please indicate how often you or other members of your household visited the lake.

Time Period	Number of Visits to the Lake
November 1999 through February 2000	
March 2000 through May 2000	
June 2000 through August 2000	
September 2000 through October 2000	
Total	

2. How many of these visits lasted longer than a single day? _______

3. In how many of the past five years did you visit the lake?_______

4. Thinking about the past year, while you were visiting Clear Lake, what percentage of your time did you spend:

Activity	Percentage
Fishing	_______%
Sailing	_______%
Recreational boating (water skiing, power boating, jet skiing, etc.)	_______%
Swimming/beach use	_______%
Nature appreciation/viewing	_______%
Snowmobiling and other winter recreation	_______%
Camping	_______%
Picnicking	_______%
Other_______________	_______%
	100%

5. During the time period of November 1999 through October 2000, how many visits to the following lakes or reservoirs did you take?

Lake or Reservoir	Number of Visits	Lake or Reservoir	Number of Visits
Lake Okoboji – East and West		Lake Odessa	
Lost Island Lake		Rathbun Reservoir	
Rice Lake		Mississippi River	
Spirit Lake		Minnesota Lakes	
Storm Lake		Wisconsin Lakes	
Tuttle Lake		Other:	
Saylorville Lake			
Lake Red Rock			
Coralville Reservoir			

6. During the course of the next year, how many trips do you expect to make to Clear Lake?

_______ trips next year

7. Suppose that the price of visiting Clear Lake increases by $10 per trip (due for example to gas prices, user fees, or equipment costs). How many times would you visit next year?

_______trips next year

8. Now suppose that the price of Clear Lake increases by $25 per trip. How many times would you visit next year?

_______ trips next year

9. On a typical visit to Clear Lake, how much money do you spend in or near the town of Clear Lake?_________

IN THE FOLLOWING SECTIONS WE WILL ASK YOU SOME QUESTIONS about potential changes to the water quality of Clear Lake during the coming years. First, however, we will give you some information on the current condition of the lake. Please read this information carefully before answering the questions that follow.

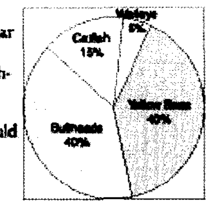
Clear Lake's Current Condition

The quality of a lake can be described in many ways. One measure of water quality is the clarity of the lake water. Water clarity is usually described in terms of how far down into the water an object is visible. The clarity of Clear Lake at the present time is about one-half to one foot. This means that objects are only visible down to about one foot under the surface of the water. The average water clarity of Clear Lake in 1953 was about ten feet.

Another measure of water quality is the amount of nutrients and other substances contained in the water. Water quality degradation can result from a number of sources, including runoff from the surrounding community containing fertilizers used for lawn care and from local agricultural sources. Currently these nutrients contribute to the occurrence of algae blooms in the lake, usually 10 to 12 times per year. Under some circumstances, these blooms can be a health concern, causing skin rashes and allergic reactions. In the past, concerns about bacteria present in Clear Lake have resulted in beach closings.

The overall quality of the water can impact on other conditions of the lake. Poor water quality results in an undesirable color and odor to the lake water. Currently, the color of Clear Lake varies between bright green and brown. The water has a mild odor that many describe as "fishy," with occasional periods of strong odor.

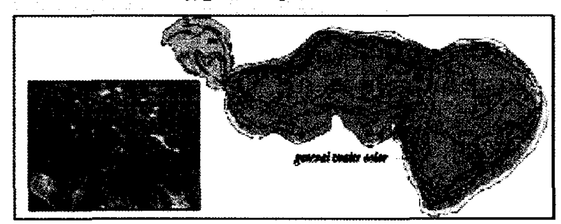
Finally, the quality of the water impacts the variety and quantity of fish in the lake. Currently, Clear Lake has a large quantity of walleye, but the largest percentages of the fish caught in the lake are fish that are considered somewhat less desirable. The chart indicates the type of fish that have been caught in the lake over the past year. While the rate at which fish are caught varies from year to year and from season to season, the typical catch rate has been 1 fish every 2 hours of fishing during the peak fishing months (May and June).



Experts believe that improved water quality would not significantly increase the number of fish in Clear Lake, but would increase the variety of fish species caught, including bass, perch, muskie, and pike.

Overall, the current condition of Clear Lake can be summarized in terms of

Water clarity	objects distinguishable 6 inches to 1 foot under water
Algae blooms	10 to 12 per year
Water color	bright green to brown
Water odor	mild odor, occasionally strong
Bacteria	possible short-term swim advisories
Fish	low diversity, good walleye

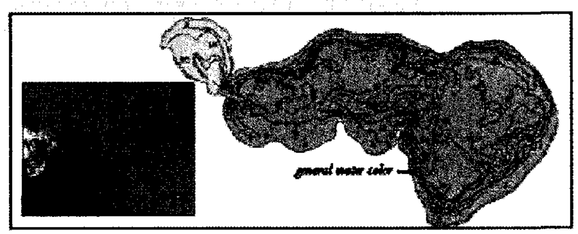


IN THIS SET OF QUESTIONS, WE WOULD LIKE TO ASK YOU about possible changes to the water quality of Clear Lake. Please answer the questions in order and do not go back and revise your earlier answers.

Plan A

If nothing is done to improve the water quality of the lake it is likely to deteriorate over the next decade. Suppose that the conditions at Clear Lake were:

Water clarity	**objects distinguishable 1 inch to 5 inches under water**
Algae blooms	**constant**
Water color	**fluorescent green**
Water odor	**always strong**
Bacteria	**frequent swim advisories and/or beach closings**
Fish	**low diversity, mostly rough fish**



10. Consider **all** of the recreation trips you made to Clear Lake in the past year. How many trips per year would you have made to Clear Lake if conditions were as described in **Plan A?** ______ trips per year.

IN THE NEXT FEW QUESTIONS, WE WILL BE ASKING YOU HOW you would vote on a special ballot regarding the water quality of Clear Lake. While there is currently no such ballot being considered, we would like you to respond *as if* you were voting on the project and, in each case, as if it were the **only** project available.

When you think about your answer, it is important to keep in mind that people tend to indicate that they would be willing to pay more money when payment is hypothetical than when they're really expected to pay. The idea is that it is very easy for people to say that they support a project when they **know** they will never have to pay any money based on their response. However, if the proposed payments are real, people may be more inclined to think about other options and what things they would have to give up to make this payment. So in answering the following questions, please keep in mind both the benefits of maintaining Clear Lake's water quality and the impact that passage of the referendum would have on your own pocketbook. In other words, please answer as if this were a **real** referendum and it was the **only** project available.

11. **Would you vote "yes" on a referendum to *maintain* the current water quality of Clear Lake and *avoid* the deteriorated water quality as described under Plan A? The proposed project would cost you $50 (payable in five $10 installments over a five year period).**

❑ NO
❑ YES

12. **To help us better understand your answers, please indicate the single most important reason for your response to the preceding question:**

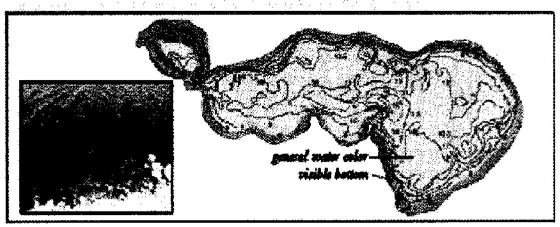
❑ **In general, avoiding Plan A is *not* a good use of my money.**
❑ **In general, avoiding Plan A is a good use of my money.**
❑ **The plan is not realistic, or unclear**
❑ **The costs of the program should be paid for by those damaging the lake, not by me.**
❑ **I already contribute to environmental causes as much as I can afford.**
❑ **No one should have the right to damage the lake in the first place.**
❑ **Other:** ______________________

Plan B

Suppose that investments could be made to actually improve the quality of Clear Lake. These investments might include establishing protection strips along the edge of the lake to reduce runoff from the surrounding area or other structural changes to the lake.

These changes would improve the lake over the next five to ten years to the following conditions:

Water clarity	**objects distinguishable 2 to 4 feet under water**
Algae blooms	**6 to 8 per year**
Water color	**green to brown**
Water odor	**occasional mild**
Bacteria	**occasional swim advisories**
Fish	**low diversity, good walleye**



13. Consider all of the recreation trips you made to Clear Lake in the past year. How many trips per year would you have made to the Lake if conditions were as described in **Plan B**? ________ trips per year

14. Would you vote "yes" on a referendum to *improve* the water quality in Clear Lake to the level described under **Plan B**? The proposed project would cost you $100 (payable in five $20 installments over a five year period).

☐ NO
☐ YES

15. To help us better understand your answers, please indicate the single most important reason for your response to the preceding question:

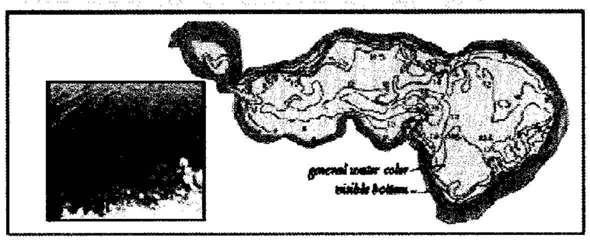
- ☐ In general, Plan B is *not* a good use of my money.
- ☐ In general, Plan B is a good use of my money.
- ☐ The plan is not realistic, or unclear
- ☐ The costs of the program should be paid for by those damaging the lake, not by me.
- ☐ I already contribute to environmental causes as much as I can afford.
- ☐ No one should have the right to damage the lake in the first place.
- ☐ Other:____________________

Plan C

Now suppose that additional investments could be made such that conditions at Clear Lake would improve further. These additional changes could include retiring land from agricultural use, and programs to control nutrient runoff from urban and agricultural lands.

Suppose these changes would improve the lake over the next ten to twenty years to the following conditions:

Water clarity	**objects distinguishable 6 to 8 feet under water**
Algae blooms	**3 to 4 per year**
Water color	**green to blue**
Water odor	**occasional mild**
Bacteria	**infrequent swim advisories**
Fish	**high diversity**



16. Consider all of the recreation trips you made to Clear Lake in the past year. How many trips per year would you have made to Clear Lake if conditions were as described in **Plan C**? ________ trips per year.

17. Would you vote "yes" on a referendum to improve the water quality in Clear Lake to the level described under **Plan C**? The proposed project would cost you $200 (payable in five $40 installments over a five year period).

- ❑ NO
- ❑ YES

18. To help us better understand your answers, please indicate the single most important reason for your response to the preceding question:

- ❑ In general, Plan C is *not* a good use of my money.
- ❑ In general, Plan C is a good use of my money.
- ❑ The plan is not realistic, or unclear
- ❑ The costs of the program should be paid for by those damaging the lake, not by me.
- ❑ I already contribute to environmental causes as much as I can afford.
- ❑ No one should have the right to damage the lake in the first place.
- ❑ Other: ______________________

IN THIS SECTION, WE WOULD LIKE TO ASK YOU ABOUT YOUR opinions regarding which lake characteristics are important to you and your views regarding some specific proposals to change Clear Lake.

19. Assume you have a total of 100 importance points to assign to the lake characteristics below. Please indicate the importance of each item by allocating your100 points among the items on this list. To indicate one item is more important to you than another, you should allocate more points to it. You do not need to give points to all of the items, but remember that the total needs to equal 100.

Water clarity	
Hard, clean, sandy lake bottom in swimming areas	
Lack of water odor	
Diversity of wildlife seen at Clear Lake	
Diversity of fish species/habitat	
Quantity of fish caught	
Safety from bacteria contamination/health advisories	
Total	100

20. A number of projects have been suggested to accomplish improvements in the lake. How do you feel about the following possibilities?

	Strongly Support	Somewhat Support	Neutral	Somewhat Oppose	Strongly Oppose
Increased park lands and recreational areas	❑	❑	❑	❑	❑
Building of a nature center or environmental park	❑	❑	❑	❑	❑
Purchase of easements for building buffer strips	❑	❑	❑	❑	❑
Increased land idling	❑	❑	❑	❑	❑
Restoration of Ventura Marsh to improve nutrient retention	❑	❑	❑	❑	❑
Non-motor boat days	❑	❑	❑	❑	❑
Increased no-wake zones	❑	❑	❑	❑	❑
Limiting motor horsepower	❑	❑	❑	❑	❑
Lake friendly restrictions on residential development	❑	❑	❑	❑	❑
Repair of storm drains	❑	❑	❑	❑	❑

21. In order to improve water quality in the lake, changes in land use in the watershed may be needed. For example, it's likely that some land will need to be changed to low-impact use. If such changes occur, which of the following land uses do you favor? Please check all that apply.

	Strongly Support	Somewhat Support	Neutral	Somewhat Oppose	Strongly Oppose
Park lands	❑	❑	❑	❑	❑
Additional Conservation Reserve Program acreage	❑	❑	❑	❑	❑
Restored woodlands	❑	❑	❑	❑	❑
Restored prairie	❑	❑	❑	❑	❑
Restored wetlands	❑	❑	❑	❑	❑
Nature conservation area	❑	❑	❑	❑	❑
Constructed ponds	❑	❑	❑	❑	❑
Hunting reserves	❑	❑	❑	❑	❑
Restored riparian zones	❑	❑	❑	❑	❑
Perennial Agriculture	❑	❑	❑	❑	❑
Other	❑	❑	❑	❑	❑

INFORMATION ON YOU AND OTHER MEMBERS OF YOUR household will help us better understand how household characteristics affect an individual's use of Clear Lake and attitudes towards changes in them. It will also help us to determine how representative our sample is of the state of Iowa. All of your answers are strictly confidential. The information will only be used to report comparisons among groups of people. We will never identify individuals of households with their responses. Please be as complete as possible.

22. Are you

❑ male ❑ female

23. What is your age?

❑ Under 18
❑ 18 – 25
❑ 26 – 34
❑ 35 – 49
❑ 50 - 59
❑ 60 - 75
❑ 76 +

24. What is the highest level of schooling that you have completed? (Please check only one)

❑ Eight years or less
❑ Some high school or less
❑ High school graduate
❑ Some college or trade/vocational school
❑ Two years of college or trade/vocational school
❑ College graduate
❑ Some graduate school
❑ Advanced degree

25. How many adults live in your household (over the age of 18)? _______

26. How many children live in your household (18 or under)?

______

27. If you are currently employed, how many hours a week do you typically work? ______

28. If you are currently employed, do you have the option of working additional hours to increase your total income?

- ❑ No
- ❑ Yes—if so, what would your hourly wage be?
  $______ per hour

29. If you answered "no" to question 28, and you could have the option of working more or less hours, which would you prefer?

- ❑ Work more hours
- ❑ Work less hours

30. What was your total household income (before taxes) in 1999?

❑ Under $10,000	❑ $40,000-$49,999
❑ $10,000-$14,999	❑ $50,000-$59,999
❑ $15,000-$19,999	❑ $60,000-$74,999
❑ $20,000-$24,999	❑ $75,000-$99,999
❑ $25,000-$29,999	❑ $100,000-$124,999
❑ $30,000-$34,999	❑ $125,000-$149,999
❑ $35,000-$39,999	❑ Over $150,000

31. Do you own a home in Clear Lake?

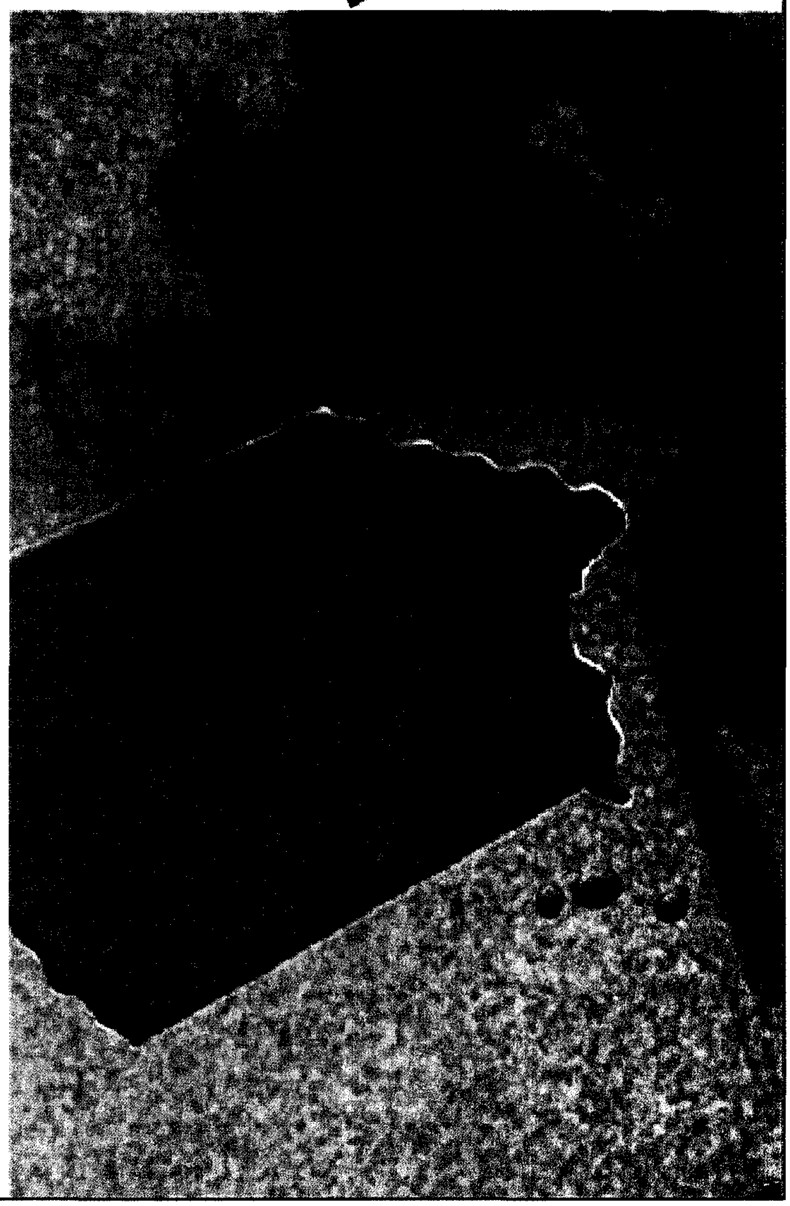
- ❑ No
- ❑ Yes, If yes, are you a year-round resident?
  - ❑ No
  - ❑ Yes

Please feel free to make any additional comments about your answers to these questions or about the survey itself. Thank you for your assistance with our Clear Lake Survey.

Comments: ______

APPENDIX 2. FIRST YEAR IOWA LAKES SURVEY

Iowa Lakes Survey 2002



Center for Agricultural and Rural Development
Resource and Environmental Policy Division

Department of Economics

IOWA STATE UNIVERSITY

In order to make sound decisions concerning the future of Iowa lakes, it is important to understand how the lakes are used, as well as what factors influence your selection of lakes to visit. The answers you give to the questions in this survey are very important. Even if you have not visited any lakes in Iowa, please complete and return the questionnaire. It is critical to understand the characteristics and views of both those who use and those who do not use the lakes.

In this first section, we would like to find out which of the lakes on the enclosed map you visited and what you did there.

1. Please indicate how often you or other members of your household visited each of the following lakes in the current and past year. Also, indicate the number of trips you anticipate making to each of the lakes in 2003. If you have not visited any lakes in Iowa, and do not plan to visit any in the upcoming year, please check this box and skip to question 2.
   - ❑ I have not, and do not plan to visit any lakes in Iowa

   If you visited lakes in Iowa that are not on this list, please count them in the "other" category at the end of the list (page 7).

		Number of visits (January-December) in:					
		2002		2001		Planned for 2003	
Name of Lake	County	Single-Day	Over-night	Single-Day	Over-night	Single-Day	Over-night
Arbor Lake	**Poweshiek**						
Arrowhead Lake	Pottawattamie						
Arrowhead Pond	**Sac**						
Avenue of the Saints Lake	Bremer						
Badger Creek Lake	**Madison**						
Badger Lake	Webster						
Beaver Lake	**Dallas**						
Beeds Lake	Franklin						
Big Creek Lake	**Polk**						
Big Spirit Lake	Dickinson						
Black Hawk Lake	**Sac**						
Blue Lake	Monona						
Bob White Lake	**Wayne**						
Briggs Woods Lake	Hamilton						
Browns Lake	**Woodbury**						
Brushy Creek Lake	Webster						
Carter Lake	**Pottawattamie**						
Casey Lake (aka Hickory Hills)	Tama						
Center Lake	**Dickinson**						
Central Park Lake	Jones						
Clear Lake	**Cerro Gordo**						
Chatfield Lake	Lee						
Cold Springs Lake	**Cass**						
Coralville Lake	Johnson						
Crawford Creek Impoundment	**Ida**						
Crystal Lake	Hancock						
Dale Maffitt Lake	**Madison**						

Name of Lake	County	Number of visits (January-December) in: 2002 Single-Day	2002 Over-night	2001 Single-Day	2001 Over-night	Planned for 2003 Single-Day	Planned for 2003 Over-night
Deer Creek Lake	Plymouth						
DeSoto Bend Lake	Harrison						
Diamond Lake	Poweshick						
Dog Creek (Lake)	O'Brien						
Don Williams Lake	Boone						
East Lake (Osceola)	Clarke						
East Okoboji Lake	Dickinson						
Easter Lake	Polk						
Eldred Sherwood Lake	Hancock						
Five Island Lake	Palo Alto						
Fogle Lake	Ringgold						
George Wyth Lake	Black Hawk						
Green Belt Lake	Black Hawk						
Green Castle Lake	Marshall						
Green Valley Lake	Union						
Hannen Lake	Benton						
Greenfield Lake	Adair						
Hawthorn Lake (aka Barnes City)	Mahaska						
Hickory Grove Lake	Story						
Hooper Area Pond	Warren						
Indian Lake	Van Buren						
Ingham Lake	Emmet						
Kent Park Lake	Johnson						
Lacey Keosauqua Park Lake	Van Buren						
Lake Ahquabi	Warren						
Lake Anita	Cass						
Lake Cornelia	Wright						
Lake Darling	Washington						
Lake Geode	Henry						
Lake Hendricks	Howard						
Lake Icaria	Adams						
Lake of the Hills	Scott						
Lake Iowa	Iowa						
Lake Keomah	Mahaska						
Lake Manawa	Pottawattamie						

		Number of visits (January-December) in:					
		2002		2001		Planned for 2003	
Name of Lake	County	Single-Day	Over-night	Single-Day	Over-night	Single-Day	Over-night
Lake McBride	Johnson						
Lake Miami	Monroe						
Lake Minnewashta	Dickinson						
Lake Orient	Adair						
Lake Pahoja	Lyon						
Lake Smith	Kossuth						
Lake Sugema	Van Buren						
Lake of Three Fires	Taylor						
Lake Wapello	Davis						
Little River	Decatur						
Little Sioux Park Lake	Woodbury						
Little Spirit Lake	Dickinson						
Little Wall Lake	Hamilton						
Littlefield Lake	Audubon						
Lost Island Lake	Palo Alto						
Lower Gar Lake	Dickinson						
Lower Pine Lake	Hardin						
Manteno Lake	Shelby						
Mariposa Lake	Jasper						
Meadow Lake	Adair						
Meyers Lake	Black Hawk						
Mill Creek (Lake)	O'Brien						
Mitchell Lake	Black Hawk						
Moorhead Lake	Ida						
Mormon Trail Lake	Adair						
Nelson Park Lake	Crawford						
Nine Eagles Lake	Decatur						
North Twin Lake	Calhoun						
Oldham Lake	Monona						
Otter Creek Lake	Tama						
Ottumwa Lagoon	Wapello						
Pierce Creed Lake	Page						

		Number of visits (January-December) in:					
		2002		2001		Planned for 2003	
Name of Lake	County	Single-Day	Over-night	Single-Day	Over-night	Single-Day	Over-night
Pleasant Creek Lake	Linn						
Poll Miller Park Lake	Lee						
Prairie Rose Lake	Shelby						
Rathbun Lake	Appanoose						
Red Haw Lake	Lucas						
Red Rock Lake	Marion						
Roberts Creek Lake	Marion						
Rock Creek Lake	Jasper						
Rodgers Park Lake	Benton						
Saylorville Dam	Polk						
Silver Lake	Delaware						
Silver Lake	Dickinson						
Silver Lake	Palo Alto						
Silver Lake	Worth						
Slip Bluff Lake	Decatur						
South Prairie Lake	Black Hawk						
Spring Lake	Greene						
Springbrook Lake	Guthrie						
Storm Lake	Buena Vista						
Swan Lake	Carroll						
Thayer Lake	Union						
Three Mile Lake	Union						
Trumbull Lake	Clay						
Tuttle Lake	Emmet						
Twelve Mile Creek Lake	Union						
Union Grove Lake	Tama						
Upper Gar Lake	Dickinson						
Upper Pine Lake	Hardin						
Viking Lake	Montgomery						
Volga Lake	Fayette						
West Okoboji Lake	Dickinson						
West Osceola	Clarke						
White Oak Lake	Mahaska						
Williamson Pond	Lucas						

		Number of visits (January-December) in:					
		2002		2001		Planned for 2003	
Name of Lake	County	Single-Day	Over-night	Single-Day	Over-night	Single-Day	Over-night
Willow Lake	**Harrison**						
Wilson Park Lake	Taylor						
Wilson Lake	**Lee**						
Windmill Lake	Taylor						
Yellow Smoke Park Lake	**Crawford**						
Yen-Ruo-Gis Lake	Keokuk						
Other Lakes in Iowa							

2. Please indicate how often you or other members of your household visited lakes or rivers in each of the following locations in the current and past year. Also, indicate the number of trips you anticipate making to each of these locations in 2003.

		Number of visits (January-December) in:					
		2002		2001		Planned for 2003	
Name of Lake		Single-Day	Over-night	Single-Day	Over-night	Single-Day	Over-night
Lakes in Illinois							
Lakes in Minnesota							
Lakes in Missouri							
Lakes in Nebraska							
Lakes in South Dakota							
Lakes in Wisconsin							
The Missouri River							
The Mississippi River							
Other Lakes and Rivers							

3. What activities do you or members of your household typically participate in during your lake visits? Check all that apply.

❑Boating ❑Jet skiing ❑Picnicking
❑Camping ❑Sailing ❑Snowmobiling and other winter recreation
❑Fishing ❑Canoeing ❑Swimming and beach use
❑Hunting ❑Other________________
❑Nature Appreciation/wildlife viewing

4. How frequently do you or your family swim in Iowa lakes?

❑Never ❑Rarely ❑Sometimes ❑Frequently

In this section, we would like to find out what features of lakes are important to you.

5. Assume you have a total of 100 importance points to assign to the following factors in **choosing a lake for recreation.** Please indicate the importance of each factor by allocating your 100 points among the items on this list. To indicate one item is more important to you than another, you should allocate more points to it. You do not need to give points to all of the items, but remember that the total needs to equal 100.

Proximity	
Water quality	
Location of friends/relatives	
Park facilities	
Activities at the lake	
Activities in the town	
Other:________	
Total	**100**

6. Again assume you have a total of 100 importance points to assign to the **lake characteristics** below. Please indicate the importance of each item by allocating your 100 points among the items on this list. To indicate one item is more important to you than another, you should allocate more points to it. You do not need to give points to all of the items, but remember that the total needs to equal 100.

Water clarity	
Hard, clean, sandy lake bottom in swimming areas	
Lack of water odor	
Diversity of wildlife	
Diversity of fish species/habitat	
Quantity of fish caught	
Safety from bacteria contamination/health advisories	
Other________	
Total	**100**

7. Which of the lakes on the list is the nearest to your permanent residence?

____

How far is this lake from where you live? _________ miles.

8. How important is the presence of the lake nearest your permanent residence (the lake you identified in question 7) to. . .

	Very Important	Somewhat Important	Neutral	Somewhat Unimportant	Very Unimportant
the economic vitality of your community?	❑	❑	❑	❑	❑
making your community an interesting and vibrant place?	❑	❑	❑	❑	❑
retaining the interest of young people to remain in your community or in attracting prospective residents to your area?	❑	❑	❑	❑	❑
area employers' ability to retain and or attract a skilled workforce?	❑	❑	❑	❑	❑
encouraging corporate decision makers to consider your area for establishing a business or expanding an existing industry?	❑	❑	❑	❑	❑

9. If the water quality of the lake you identified in question 7 were significantly improved, how important do you think the lake could be to. . .

	Very Important	Somewhat Important	Neutral	Somewhat Unimportant	Very Unimportant
the economic vitality of your community?	❑	❑	❑	❑	❑
making your community an interesting and vibrant place?	❑	❑	❑	❑	❑
retaining the interest of young people to remain in your community or in attracting prospective residents to your area?	❑	❑	❑	❑	❑
area employers' ability to retain and or attract a skilled workforce?	❑	❑	❑	❑	❑
encouraging corporate decision makers to consider your area for establishing a business or expanding an existing industry?	❑	❑	❑	❑	❑

Information on you and other members of your household will help us better understand how household characteristics affect an individual's use of Iowa lakes and attitudes towards changes in them. It will also help us to determine how representative our sample is of the state of Iowa. All of your answers are strictly confidential. The information will only be used to report comparisons among groups of people. We will never identify individuals or households with their responses. Please be as complete as possible. Thank you.

10. What is your age?

❑Under 18	❑ 50 - 59
❑18 – 25	❑60 - 75
❑26 – 34	❑76 +
❑35 – 49	

11. Are you

❑male ❑ female

12. What is the highest level of schooling that you have completed? (Please check only one)

❑Some high school or less
❑High school graduate
❑Some college or trade/vocational school
❑College graduate
❑Advanced degree

13. How many adults (including yourself) live in your household? _______

14. How many children live in your household (18 or under)?________

15. If you are currently employed, how many hours a week do you typically work? _______

16. If you are currently employed, do you have the option of working additional hours to increase your total income?

❑no

❑yes—if so, what would your hourly wage be?

$_______per hour

17. If you answered "no" to question 16, and you could have the option of working more or less hours, which would you prefer?

❑Work more hours ❑Work the same number of hours

❑Work less hours

18. What is your total household income (before taxes) for 2002?

❑ Under $10,000	❑ $40,000-$49,999
❑ $10,000-$14,999	❑ $50,000-$59,999
❑ $15,000-$19,999	❑ $60,000-$74,999
❑ $20,000-$24,999	❑ $75,000-$99,999
❑ $25,000-$29,999	❑ $100,000-$124,999
❑ $30,000-$34,999	❑ $125,000-$149,999
❑ $35,000-$39,999	❑ Over $150,000

19. Do you own a home on a lake in Iowa?
❑no

❑yes, If yes, are you a year-round resident?
❑ yes
❑ no

20. Do you own a home on a lake outside of Iowa?
❑ yes ❑no

21. Do you belong to a lake protection association?
❑yes ❑ no

22. Are you an area employer?
❑yes ❑no

23. Are you involved with community development efforts and/or with making decisions that impact the entire community (for example Chamber of Commerce, Jaycees, etc.)?
❑ yes ❑no